The TB Structural Genomics Consortium: Providing a Structural Foundation for Drug Discovery

Celia W. Goulding[3], Marcin Apostol[3], Daniel H. Anderson[3], Harindarpal S. Gill[3], Clare V. Smith[4], Mack R. Kuo[4], Jin KukYang[8], Geoffrey S. Waldo[1], Se Won Suh[8] , Radha Chauhan[9,10], Avinash Kale[10], Nandita Bachhawat[9] and Shekhar C. Mande[10], Jodie M. Johnston[11], J. Shaun Lott[11], Edward N. Baker[11], Vickery L. Arcus[11], David Leys[12], Kirsty J. McLean[12], Andrew W. Munro[12], Joel Berendzen[2], Vivek Sharma[4], Min S. Park[1], David Eisenberg[3], James Sacchettini[4], Tom Alber[5], Bernhard Rupp[6], William Jacobs, Jr.[7] and Thomas C. Terwilliger[1,*]

[1]Bioscience Division, Mail Stop M888 Los Alamos National Laboratory, Los Alamos, NM 87545 USA; [2]Physics Division, Mail Stop D454, Los Alamos National Laboratory, Los Alamos, NM 87545 USA; [3]Howard Hughes Medical Institute, Molecular Biology Institute, UCLA- DOE Lab of Structural Biology, UCLA Box 951570, University of California, Los Angeles, California 90095-1570; [4]Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843-2128; [5]University of California, Berkeley, CA 94720; [6]Lawrence Livermore National Laboratory, Livermore, CA 94551; [7]Yeshiva University Albert Einstein School of Medicine, Bronx, NY 10461; [8]Structural Proteomics Laboratory, School of Chemistry and Molecular Engineering, Seoul National University, Seoul 151-742, Korea; [9]Institute of Microbial Technology, Sector 39-A, Chandigarh; [10]Centre for DNA Fingerprinting and Diagnostics, ECIL Road, Nacharam, Hyderabad, India; [11]School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand and [12]Department of Biochemistry, University of Leicester, The Adrian Building, University Road, Leicester, LE1 7RH, UK

Abstract: Structural genomics, the large-scale determination of protein structures, promises to provide a broad structural foundation for drug discovery. The tuberculosis (TB) Structural Genomics Consortium is devoted to encouraging, coordinating, and facilitating the determination of structures of proteins from *Mycobacterium tuberculosis* and hopes to determine 400 TB protein structures over 5 years. The Consortium has determined structures of 28 proteins from TB to date. These protein structures are already providing a basis for drug discovery efforts.

INTRODUCTION

Tuberculosis (TB) is caused by the bacterial pathogen *Mycobacterium tuberculosis* (*Mtb*), and kills 2-3 million people around the world each year, more than any other infectious disease. The rise in TB incidence over the two last decades is due partly to TB deaths in HIV-infected patients and partly to the emergence of multidrug resistant strains of the bacteria. This rapid increase in the disease has led to the World Health Organization funding a large effort towards stopping this disease before it becomes a global epidemic. The tuberculosis (TB) Structural Genomics Consortium was founded in the fall of 2000 with the goal of encouraging, coordinating, and facilitating the determination of structures of proteins from *Mycobacterium tuberculosis*. The Consortium aims to determine 400 TB protein structures over 5 years.

*Address correspondence to this author at the Bioscience Division, Mail Stop M888 Los Alamos National Laboratory, Los Alamos, NM 87545 USA; E-mail: terwilliger@LANL.gov

This review briefly describes the Consortium and how it is designed to help its members collectively determine more structures than they would working independently. We also summarize some of the initial structural results on TB proteins and their implications for TB drug discovery. Goulding *et al.* of the Eisenberg laboratory at UCLA describe the structures of glutamine synthetase, antigen 85B, a predicted disulfide isomerase protein, a secreted protein of unknown function (Rv1926c) and dihydroneopterin aldolase (FolB). The Sacchettini laboratory at Texas A&M University describes structures of mycolic acid methyltransferases (Smith & Sacchettini), an enoyl reductase involved in fatty acid biosynthesis (Kuo & Sacchettini) and enzymes from glyoxylate shunt pathway (Sharma & Sacchettini). The Suh laboratory at Seoul National University describes efforts to use the structure of Rv2002 to identify the function of this protein (Yang *et al*). The Mande laboratory at the Centre for DNA Fingerprinting and Diagnostics describes its studies on inositol 1-phosphate synthase and alkylhydroperoxidase C (Chauhan *et al*). The Baker laboratory (Johnston *et al*) describes a case of mistaken identity (menG) clarified by a protein structure. Finally, the Munro laboratory at the University of Leicester

1568-0053/02 $35.00+.00

describes the structures of cytochrome P450 enzymes (Leys *et al*).

THE CONSORTIUM APPROACH TO STRUCTURAL GENOMICS AND DRUG DISCOVERY

Structural genomics is the large-scale determination and analysis of protein structures. It is a new field that has been made possible by major technological improvements in structure determination and by the many completed genomes revealed by the genome projects [1, 2, 3]. One very promising application of large-scale structure determination is to provide a framework for drug discovery on a genomic scale. The TB Structural Genomics Consortium was created for this purpose. A premise of this project is that determining the structures of many of the proteins from TB facilitates drug discovery efforts using high-throughput screening, genetic identifications of key TB genes, and structure-based approaches. Some of the members of the Consortium had been carrying out structure determinations of proteins from TB before the start of the Consortium, and others have only begun more recently.

There are several major advantages to our consortium-based approach to determining structures of many proteins from a single organism. The first, a rather simple but important advantage, is that the members of the Consortium can communicate with each other easily about what proteins each is working on. This helps avoid duplication and encourages collaboration. The second advantage is that members can share the tools, often specific for producing TB proteins, that they develop, and can share information on which methods are working and which are not. A third advantage is that the Consortium has centralized facilities to carry out much of the standardized work in an efficient way for the Consortium as a whole. The TB Structural Genomics Consortium has central facilities for high-throughput cloning and expression testing, crystallization, and X-ray data collection. Individuals in the Consortium work on proteins that interest them. In parallel, the TB genes targeted by all members of the Consortium are cloned *en masse* in the central facilities at Los Alamos National Laboratory. The clones are tested for expression and the results are available to the whole Consortium. The Los Alamos protein production facility and individuals in the Consortium can send proteins to the crystallization facility at Lawrence Livermore National Laboratory and have X-ray data collected at the facilities of the Consortium at Brookhaven National Laboratory.

The TB Structural Genomics Consortium has over 200 members in 13 countries around the world. The Consortium is open and any researchers actively working on structural genomics of TB may join. The membership and publicly available data for the Consortium can be found at http://www.doe-mbi.ucla.edu/TB. To date, the Consortium determined some 28 structures of TB proteins (including some that had been determined just before the Consortium was formed). Many of these protein structures and their analyses that are providing encouraging information relevant to TB drug discovery are described in the following sections.

1. Recent Structures of *Mycobacterium tuberculosis* Proteins and their Relevance as Potential Drug Targets (Celia W. Goulding, Marcin Apostol, Daniel H. Anderson, Harindarpal S. Gill & David Eisenberg)

1.1. Identifying Potential Drug Targets

Several computational methods have been developed which can predict the biological roles of proteins by analyzing functional relationships among proteins rather than sequence similarities. Three of these methods are the phylogenetic profile [4] , domain fusion [5] and gene clustering [6, 7] methods. These methods are described in a comprehensive review [8]. Utilizing these methods [9] , potential drug targets were identified as proteins functionally linked to known targets of anti-TB drugs [10] and to proteins known to be essential for bacterial survival [11]. Inhibiting these functionally related proteins should have a similar effect on the organism as inhibiting the present drug targets, since the same processes or pathways would be disrupted.

Other proteins that might be good targets for anti-TB therapy include extracellular proteins that are involved in virulence, and persistence determinants [12] , since the *Mtb* cell envelope is impermeable to many antibacterial agents. *Mtb* secretes many proteins, some of which have been characterized as antigens. Potential therapeutic targets also include proteins involved in iron acquisition, which is a critical process for *Mtb* survival [13]. These proteins include iron-regulatory proteins and enzymes involved in the production of secreted, iron siderophores such as exochelins and mycobactins. Proteins specific to mycobacteria also afford possible targets, since these may provide adaptations exclusive to the virulence and pathogenecity of mycobacteria. In particular, 10% of the *Mtb* genome consists of genes that encode *Mtb*-specific PE, PPE and PE-PGRS proteins [14]. The PE family is named after the presence of the motif proline-gluamic acid (PE) at positions 8 and 9 in a highly conserved N-terminal domain of approximately 110 amino acids, the C-terminal region varies drastically within the family. The PPE family resembles the PE family with a highly conserved N-terminal region (~180 amino acids), containing the motif proline-proline-glutamic acid at positions 7-9. Lastly, the PE-PGRS (polymorphic GC-rich sequence) family represents an extension of the PE protein family with multiple repeats of glycine-glycine-alanine (asparagine) motifs. The function or functions of these proteins are unknown though they have been implicated in virulence [15].

To reduce the difficulty of structure determination, we imposed several additional criteria on our choice of *Mtb* proteins. The proteins must contain no predicted transmembrane helices, must be smaller than 50,000 Daltons, and all hypothetical proteins must have homologs in three different organisms. Finally, to strengthen our drug target strategy, we imposed the restraint that *Mtb* drug targets must not have essential mammalian homologs.

1.2. Mtb Protein Crystal Structures Solved at UCLA

Two out of the five structures were solved by multiwavelength anomalous diffraction from

selenomethionine protein derivatives, the remaining three were solved by molecular replacement. The cloning, expression, and structure determination of proteins Rv2878c, Rv3607c, Rv1926c have been previously reviewed [16]. The solution for Rv2220 is discussed in [17], and the solution for Rv1886c is discussed in [18].

Glutamine Synthetase-- Rv2220

The product of the *glnA1*gene from *Mtb*, glutamine synthetase (TB-GS), is extracellularly released during the early stages of infection [19, 20]. TB-GS is thought to be necessary for the synthesis of poly- (L-glutamine-L-glutamate) chains 21] , a constituent unique to pathogenic mycobacterial cell walls, comprising 10% of bacterial mass [22]. These chains have been reported to be tightly associated with the peptidoglycan layer of the cell wall and therefore may play an important role in *Mtb* virulence. In parallel with the structure solution of TB-GS [17, 23] , the laboratory of Marcus Horwitz at UCLA has demonstrated that two known bacterial GS inhibitors, L-methionine-S-sulfoximine and phosphinothricin, selectively block the growth or normal cell wall development of *Mtb* by inhibiting extracellular TB-GS molecules [19] , suggesting that TB-GS is an attractive target for therapy.

The structures of TB-GS [17, 23] and *Salmonella typhimurium* GS [24] have been useful in elucidating the mechanism of inhibition of GS by small molecules phosphinothricin and L-methionine-S-sulfoximine. TB-GS and *S. typhimurium* GS share a 52% amino-acid sequence identity. The *glnA*gene isolated from each organism was cloned into an *E. coli* expression vector and the functionality of the products was demonstrated by their ability to restore growth to an *E. coli* glutamine auxotroph. Although the proteins were purified under similar conditions, TB-GS crystallized in distinct conditions from that of *S. typhimurium* GS. Figure (**1a**) shows a model of the *S. typhimurium* GS-adenosine triphosphate (ATP)-phosphinothricin structure. The structure of TB-GS will be described elsewhere [23]. Together the models illuminate the characteristics of potent GS inhibitors. They illustrate that several flexible loops containing key catalytic residues close over each active site and shield an intermediate from solvent during the course of glutamine synthesis. These same loops make strong electrostatic contacts with bound inhibitors that mimic the intermediate state of the enzyme in glutamine synthesis, trapping the inhibitors in the active site cavity. The GS models provide a basis for future work in computational drug-design studies.

Ag85B - Rv1886c

M. tuberculosis expresses three closely related mycolyl tranferases, also known as antigen 85 proteins [ag85A, ag85C (32 kDa each); ag85B (30 kDa)]. The three mycolyl transfereases contribute to cell wall synthesis by catalyzing transfer of mycolic acid from one trehalose 6-monomycolate to another, resulting in trehalose 6,6'-dimycolate and free trehalose [25]. Mycolic acids are long-chain, α-alkyl β-hydroxyl fatty acids that form a major component of the *Mtb* cell wall. The 30 kDa major secreted protein (ag85B) is the most abundant protein exported by *Mtb*, and is a potent immunoprotective antigen and a leading drug target [26]. Due to its location at the cell wall and its involvement in cell wall biogenesis, the ag85B protein maybe a relatively accessible drug target (as is glutamine synthetase, described earlier). The inhibition of ag85B may block the synthesis of

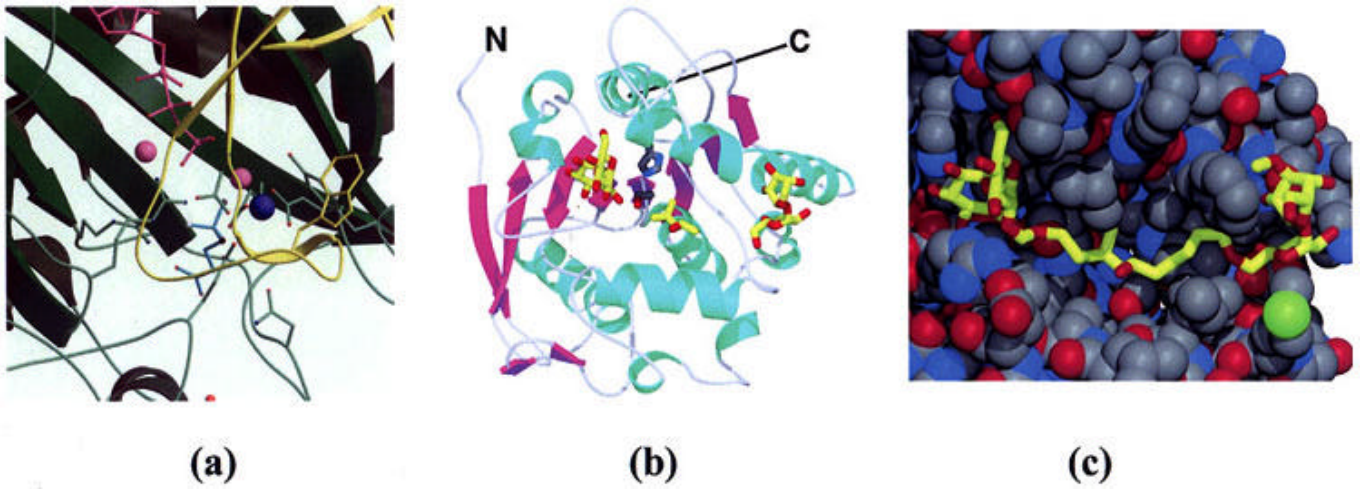

Fig. (1). The active site of glutamine synthetase with inhibitor phosphinothricin bound and three-dimensional structure of *Mtb* antigen 85B. (a) A model of phosphinothricin bound in the active site of *Salmonella typhimurium* GS is shown. The elongated pink molecule is ATP; the pink spheres are manganese ions; the light blue molecule is phosphinothricin. The dark blue sphere marks the ammonium binding site. The figure was generated with the program Molscript. (b) Ribbon diagram of trehalose-bound ag85B, showing its α/βhydrolase fold. The structure is color coded by secondary structure. Cyan ribbons represent α-helices, magenta ribbons β-strands, and coil and turn regions are white. The N and C termini are indicated in the illustration. Active site Ser126 and His262 CA and side-chain atoms are shown (protein C atoms are gray; N, blue; O, red). Two trehalose and the central MPD molecules are shown with yellow for C atoms and red for O atoms. This figure was produced using XtalView, and were rendered with Raster3D. (c) The proposed class of drugs consists of two trehalose molecules with amphiphilic linkers. The proposed drug has been modeled connecting the two trehalose molecules with a linker where the MPD residues in Fig 1b.

the cell wall and therefore inhibit growth of *Mtb*. Mycolyl transferase activity is unique to mycobacteria, making ag85B a very good anti-TB drug target.

The structure of ag85C had been previously solved [27] , and as expected considering a 73 % amino acid sequence identity, the ag85B and ag85C structures are similar. The structure of ag85B (Fig **1 b**.; [18] consists of a single-domain monomeric protein that adopts an α/β-hydrolase fold [28]. The central 8-stranded β-sheet is surrounded on both sides by α-helices. The sheet is formed in such a way that the outer two strands, β1 and β8, are oriented perpendicular to each other, resulting in a left-handed superhelical twist. The first two strands, β1 and β2, are antiparallel and connected by a hairpin turn. The remainder of the sheet is composed of ?-strands running parallel to strand β1, with the strands being connected by either a single, relatively long α-helix, or a series of smaller α-helices. Interestingly, ag85B binds two trehalose molecules and a 2-methyl-2,4-pentanediol (MPD) molecule. It has been proposed that MPD could mimic the partially hydrophilic head of the intermediate mycolate ester. Close to these active sites is a hydrophobic groove that could be the binding site of the mycolate β-chain.

Anderson *et al.* [18] proposed that a class of groove-binding inhibitors of the antigens 85 could be constructed by linking two trehalose molecules by an appropriately designed amphiphilic chain (Fig. **1c**). An inhibitor has been designed based on this idea but has been shown to have little influence on *Mtb* growth (Harth and Anderson, unpublished data). Further inhibitor designs are being pursued to obtain a molecule which will bind tightly to the active site groove.

Disulfide bond isomerase - Rv2878c

Disulfide bonds are vital for protein folding and stability. Rv2878c is a secreted protein with sequence homology to disulfide bond isomerase (Dsb) proteins such as *E. coli*DsbE. In gram-negative bacteria, disulfide bonds are usually formed in the periplasmic space by protein disulfide oxidoreductases [29]. Experimental evidence indicates that *E. coli*DsbE is a weak reductant that is involved in cyctochrome c maturation in the periplasm [30]. However, since gram-positive bacteria do not have a periplasmic space, the disulfide oxidoreductases are secreted into the extracellular space surrounding the cell [31, 32]. *Mtb* has two Dsb-like proteins, both of which are secreted (Rv2878c and Rv1677), and one predicted integral membrane protein (Rv2874), which is thought to complete the redox cycles of the two secreted Dsb-like proteins.

The crystal structure of *Mtb* DsbE (Fig. **2a**) reveals one domain that contains a thioredoxin fold [33] , with a distinct structural motif consisting of a four-stranded β-sheet corresponding to β3,β4,β6 and β7 and three flanking α-helices corresponding to α3, α5 and α6 [34]. In addition to the thioredoxin fold, a short α-helix (α1), two β-strands (β1 and β2) and another short α-helix (α2) appear at the N-terminus, and a long α-helix (α4) and a β-strand (β5) are found after the β3-α3-β4 motif of the thioredoxin fold. The cysteines of the active site Cys-X-X-Cys are in their reduced state (Fig. **2b**). The precise function of *Mtb* DsbE is unknown, although preliminary experimental evidence indicates that *Mtb* DsbE is an oxidant rather than a weak reductant, as is *E. coli*DsbE [30]. Therefore one could speculate that *Mtb* DsbE could potentially protect mycobacteria from oxidative damage by macrophages and/or correct incorrectly formed disulfide bonds in secreted proteins. It may be possible to design an inhibitor of *Mtb* DsbE that would react with the free thiols of the active site to inhibit the disulfide bond isomerase/redox activity of *Mtb* DsbE [35].

Fol B - Rv3607c

Folate derivatives are essential cofactors in the biosynthesis of purines, pyrimidines and amino acids. Mammalian cells are able to utilize pre-formed folates after uptake by a carrier-mediated active transport system. Most microbes, such as *Mtb*, lack this uptake system and must synthesize folates *de novo* from guanosine triphosphate [36]. Two enzymes from this pathway, dihydropteroate synthase and dihydrofolate reductase have been used as drug targets for antibacterial chemotherapy [37]. It has been shown that sulfonamides such as trimethoprim and epiroprim inhibit the activity of these enzymes, but are poor overall inhibitors of mycobacteria [38, 39]. Therefore, we have chosen another enzyme, 7,8-dihydroneopterin aldolase, in the biosynthetic pathway of folic acid to explore the possibility of an inhibitor.

Dihydroneopterin aldolase (FolB) is an enzyme in the folic acid biosynthetic pathway. It catalyzes the conversion of 7,8-dihydroneopterin to 6-hydroxymethyl-7,8-dihydro-pterin. The structure of *Mtb* FolB (Fig. **3a**) is extremely similar to that of FolB from *S. aureus* [40] and FolX from *E. coli* [41]. The structure is an octamer, consiting of a dimer of tetramers that appears to have a pterin compound bound at the active site [42]. Each monomer consists of a sequential four-stranded antiparallel β-sheet. Layered on one side of the β-sheet is a short α-helix, and two long antiparallel α-helices are inserted in a sequence segment between β-strands 2 and 3. Between strands 1 and 2, there is an insert containing a short α-helix that is situated on the same face of the β-sheet as the other helices. Four ellipsoidal monomers come together to form a 16-stranded anti-parallel β-barrel, surrounded by a ring of α-helices. The catalytic site is on the interface of two of the monomers.

Comparisons of *Mtb* FolB with the *S. aureus* FolB and *E. coli* FolX structures show that the main difference is in the electrostatic surface potential of the proteins. The latter two structures have no distinct surface charge, but *Mtb* FolB has an overall negative charge on its surface (Fig. **3b**). Also, the pore of the *S. aureus* FolB structure (Fig. **3c**) is more positively charged than that of *Mtb* FolB (Fig. **3b**). With this in mind, it may be possible to design an inhibitor specific to TB FolB consisting of a cationic pterin derivative.

Secreted protein - Rv1926c

Rv1926c is a major secreted protein of unknown function, specific to mycobacteria. It has been shown to

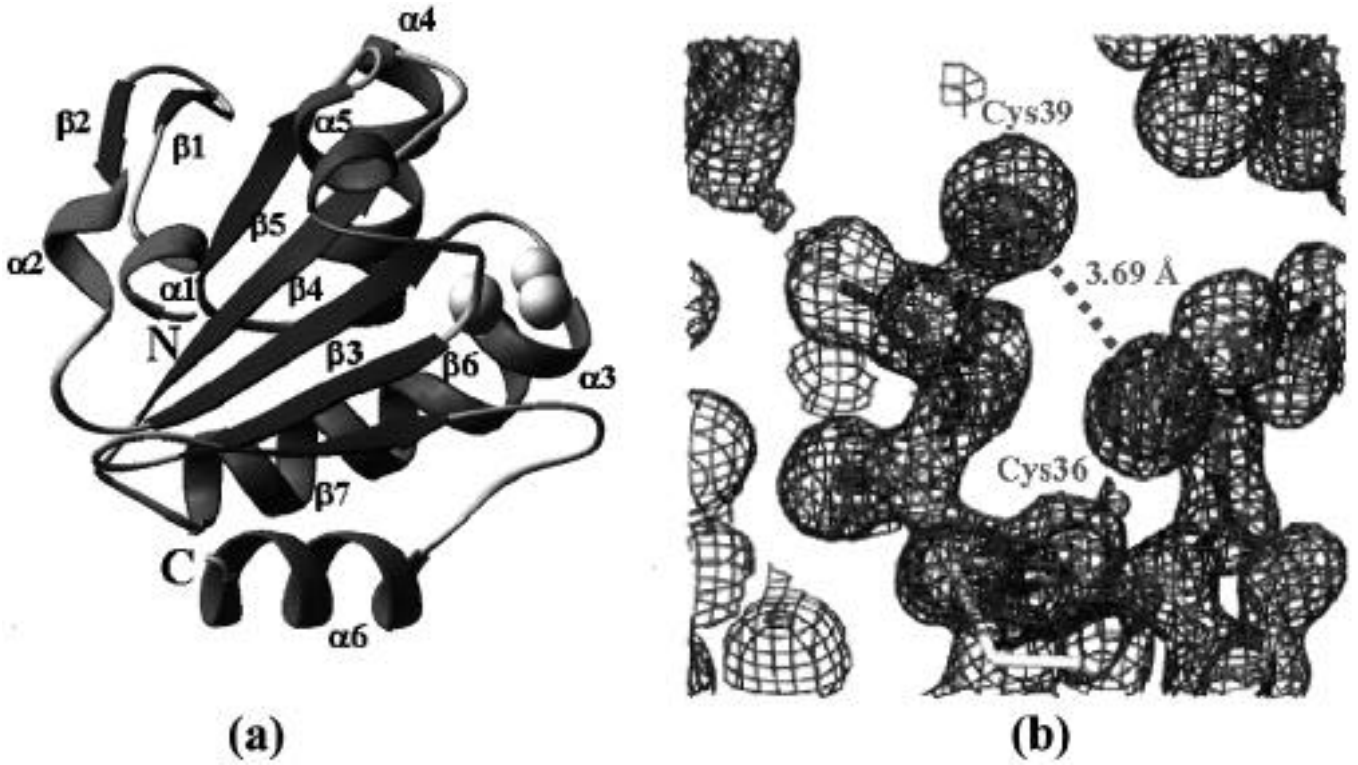

Fig. (2). Three-dimensional structure of *Mtb* DsbE. (a) Ribbon diagram of the monomer. The active site sulfur and β-carbon atoms are shown in light grey spheres. This figure was generated using RIBBONS. (b) The *Mtb* DsbE active site, a view of the final 2Fo-Fc electron density map at the active site region. The electron density is contoured at 1. 2σ. The distance between the two sulfur atoms is 3. 69 Ångstroms. The figure was generated using RIBBONS.

stimulate humoral immune responses in guinea pigs infected with virulent *Mtb* [43]. T-cell epitope mapping showed that Rv1926c contained a highly immunodominant region at its N-terminus, though this may be the region of the signal peptide [44]. Using a Tn*552'pho*A *in vitro* transposition system, Braunstein *et al.* [45] confirmed that Rv1926c is a secreted protein, and have shown it to be a cell-envelope-associated protein that may participate in virulence. The structure of Rv1926c is a β-sandwich fold consisting of two anti-parallel sheets, except for the N-terminal β-strand which forms a parallel β-sheet with the C-terminal β-strand (Fig. **3d**) [46].

The fold of Rv1926c places it in the immunoglobulin superfamily [47]. Rv1926c has structural similarity to cell-surface binding proteins such as arrestin [48] , adaptin [49] and invasin [50]. Other proteins with a high degree of structural similarity include several that bind carbohydrates, Ca2+ ions, lipids and cholesterol (which was shown to be essential in entry of mycobacteria into macrophages [51]. The electrostatic surface potential of Rv1926c reveals a negatively charged pocket and a positively charged channel that could possibly bind a peptide or one of the above co-factors.

Structural similarity implicates Rv1926c in possible host-bacterial interactions for endocytosis or phagocytosis. Further work will include genetic knock-out studies of the Rv1926c gene in *Mtb* to test for its role in infection.

2. Structures of the Cyclopropane Synthases from *M. tuberculosis* (CmaA1: Rv3392c, CmaA2: Rv0503c and PcaA: Rv0470c) (Clare V. Smith and James C. Sacchettini)

The genome sequence of *Mtb* revealed a family of highly homologous genes sharing 50-75% identity and thought to be *S*-adenosyl-L-methionine dependent methyltransferases (SAM-MTases) [14]. At least four members of this family *cmaA1* (Rv3392c), *cmaA2* (Rv0503c), *pcaA* (Rv0470c) and *mmaA2* (Rv0644c) had been previously reported to encode cyclopropane synthases [52, 53, 54, 55]. This designation is based on their similarity to the *E. coli* enzyme cyclopropane fatty acid synthase (CFA synthase), which catalyzes the addition of a methyl group to a double bond on an acyl chain of the phospholipid bilayer [56]. In mycobacteria, cyclopropane synthases lead to the addition of a cyclopropane ring on mycolic acids, which are in turn critical to the structure and function of the cell envelope [52, 57, 53, 55].

There are three classes of mycolic acids synthesized in *Mtb*. These are the alpha-, keto- and methoxymycolates, which are classified according to their modifications at the proximal and distal positions (Fig. **4a**). The large number of cyclopropane synthases in the genome of *Mtb* suggests that each member of the gene family may have a very specific role in catalyzing the reaction on one particular mycolic acid at either the distal or proximal position.

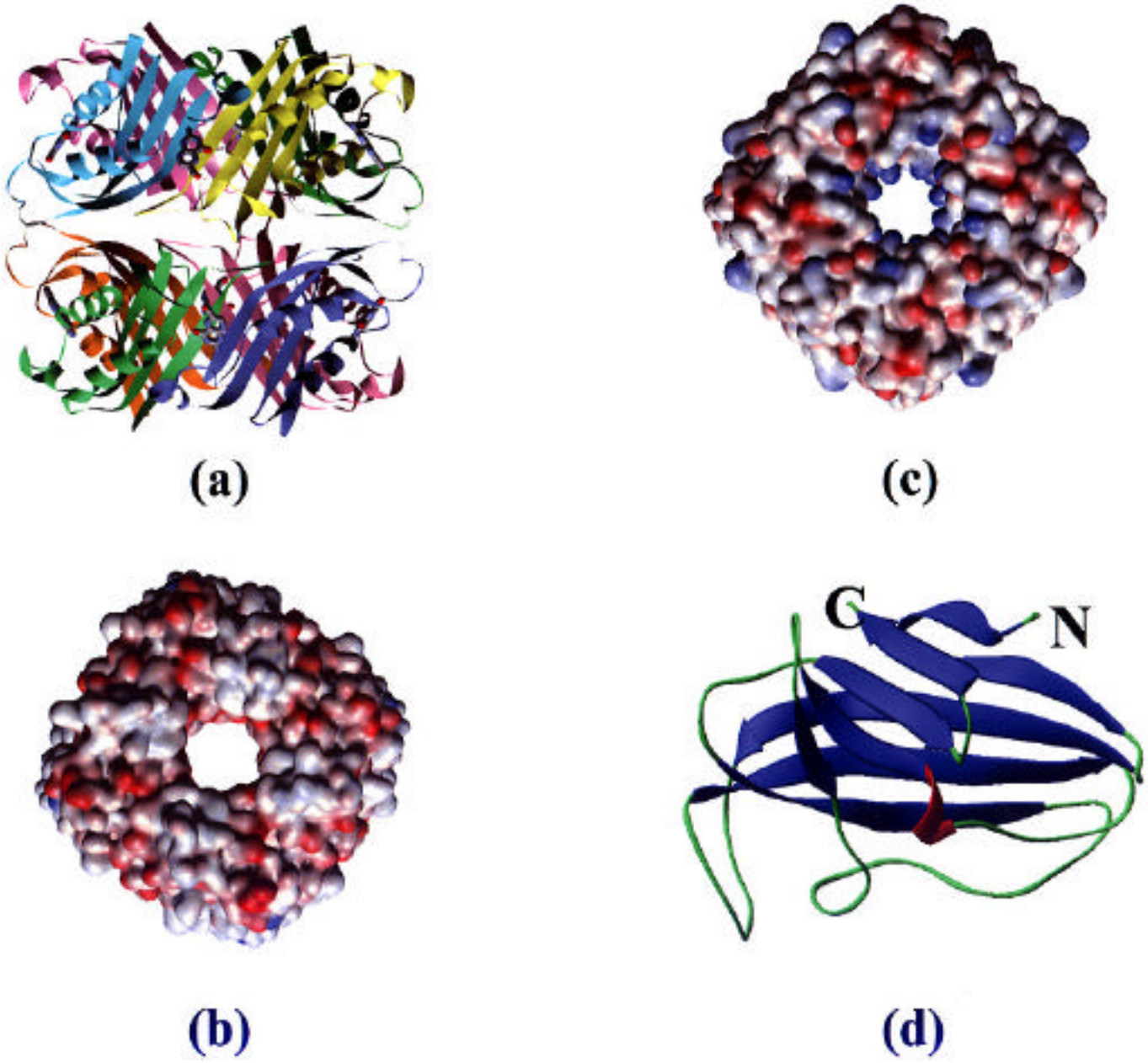

Fig. (3). Three-dimensional structure of *Mtb* FolB and three dimensional structure of Rv1926c. (a) The ribbon diagram of the structure of FolB is an octomer, which consists of a dimer of tetramers that appears to have a pterin compound bound to its active site which is on the interface of two of the monomers. (b) and (c) The electrostatic surface potential of *Mtb* FolB (b) is more negatively charged than that of *S. aureus* FolB (c). The ribbons diagrams and electrostatic surface potentials were produced with WEBLAB VIEWER PRO 3. 7. (d) The ribbon diagram of the structure of Rv1926c resembles an immunoglobulin fold with the N-terminal and C-terminal β-strands forming a parallel β-sheet.

Although the precise role of the cyclopropane ring has not been defined, there are several lines of evidence suggesting that it is important in bacterial pathogenesis. Cyclopropanation is a modification of mycolic acids which is associated with slow-growing pathogenic strains such as *M. tuberculosis* and *M. leprae*, but is not common in the cell wall of fast growing non-pathogenic strains such as *M. smegmatis* [58]. Studies in mycobacteria have suggested that cyclopropane ring modification may increase the resistance to oxidative stress [55] and have an effect on the fluidity and permeability of the cell wall [52, 59, 60]. Recently Glickman *et al.* [53] have identified *Mtb pcaA* as a gene encoding a cyclopropane synthase involved in cyclopropanation at the proximal position of alpha mycolates. Further characterization of the mutant showed that *pcaA* was required for long-term mycobacterial persistence and virulence in an animal model of infection [53].

To date, crystal structures of three cyclopropane synthases from *M. tuberculosis* have been solved, including that of PcaA, the member already identified as a persistence factor [61]. PcaA is believed to catalyze the transfer of a

methyl group to a *cis*double bond at the proximal position of alpha mycolates. CmaA2 also acts at the proximal position but converts a *trans*double bond to a cyclopropane ring on keto- and methoxymycolates [52, 57]. The third enzyme for which the structure has been solved is CmaA1. This enzyme is thought to act at the distal position giving rise to the cyclopropane ring of alpha mycolates [55] (Fig. **4b**).

The basic structures for the three cyclopropane synthases CmaA1, CmaA2 and PcaA are very similar (Fig. **4c**). The core region contains a seven stranded β-sheet which is all parallel apart from β7 which is antiparallel. The α-helices flank each side of the sheet and run in the same N- to C-orientation. In addition, two long α-helices lie adjacent to the C-terminal ends of the β-sheet, enclosing the SAM/SAH (S-adenosyl - L-methionine / S-adenosyl - L - homocysteine) cofactor binding site. The structures of the cyclopropane synthases are similar to other SAM-MTases in the protein database, including DNA methyltransferases and catechol-O-methyltransferases (Fig. **4d**). Overall, the sequence similarity to these other methyltransferases is low, but the motifs which make up the cofactor binding site are conserved in sequence and structurally [62, 63, 64, 65].

One of the most interesting features of the cyclopropane synthases to be revealed by their structures is a tunnel approximately 15 Å by 10 Å wide extending from the surface of the protein to the cofactor binding site. This tunnel is lined primarily by the side chains of hydrophobic residues that are highly conserved across the three enzymes. Structures of CmaA1 and CmaA2 in complex with SAH and the lipid-like detergents cetyltrimethylammonium bromide (CTAB) or didecyldimethylammonium bromide (DDDMAB) indicate that this is the binding site for the acyl substrate. The co-crystal structure with lipid demonstrates important interactions between the protein and the acyl chain providing us with insights into the way in which the reactive group may sit in the active site and the length of the acyl chain that these enzymes may accept.

Analysis of a group of enzymes such as the cyclopropane synthases using a structural genomics approach provides high resolution insights into such details as the active site architecture. An unexpected finding was a bicarbonate ion revealed in the active site of all three enzymes. One conclusion which can be drawn based on the similarity of the three cyclopropane synthases in their cofactor binding site and active sites is that these enzymes are likely to operate via the same catalytic mechanism. Although there are still many questions about the mechanism of cyclopropane synthases, the structures give us many clues and help us direct our questions.

What is the basis for substrate specificity of CmaA1, CmaA2 and PcaA? Biological data suggests that CmaA1, CmaA2 and PcaA have non overlapping functions *in vivo*. Comparison of the structures of CmaA2 and PcaA, two enzymes that act at the proximal position producing *trans* and *cis* cyclopropane rings respectively, reveals only minor differences between the two structures. Further experiments are undoubtedly needed before we understand fully the basis for the *cis* versus *trans* substrate specificity.

Although the point at which these cyclopropane synthases act in mycolic acid biosynthesis is not known, it is believed that they act on a long chain meromycolate intermediate that is linked to acyl carrier protein (AcpM). AcpM facilitates processing of intermediates between the dissociated enzymes of FAS II. Superimposition of the PcaA, CmaA2 and CmaA1 structures demonstrates that there is one region where the proximal versus distal-acting enzymes differ. This region is shifted in the distal cyclopropanating enzyme by approximately 5 Å (Fig. **4e**). Examination of the molecular surfaces in this region suggests that there is a conserved basic/hydrophobic patch. This molecular surface architecture is also seen in holo- (acyl carrier protein) synthase (AcpS) [66] and β-ketoacyl-ACP synthase III (FabH) [67]. It seems that this maybe a common feature of enzymes of fatty acid biosynthesis and a likely site for AcpM interaction. If this is indeed the region where AcpM interacts with the cyclopropane synthases, then acyl-AcpM may sit closer to the active site in the case of PcaA and CmaA2, so favoring a reaction at the proximal position. In contrast, acyl-AcpM would bind further from the active site in CmaA1, so favoring a reaction at the distal position of the mycolates, suggesting a molecular ruler basis for substrate specificity [61].

One feature that would be desirable in future antimycobacterial compounds is potency against bacteria in the persistent phase of infection. This would shorten the necessary duration of chemotherapy. PcaA has been shown to be necessary for a chronic lethal TB infection and experiments are underway to test this requirement for the other members of this family. The similarity of this family of enzymes in their active site suggests the possibility that one inhibitor may well be effective against a number of targets, reducing the chance for the development of drug resistance. Because mammals lack this enzyme activity, a homologous enzyme will not be present in humans, so the enzymes are likely to be very good drug targets. The high resolution structural data, together with biochemical assays, can now form the basis for the development of new inhibitors active against the cyclopropane synthases.

3. Glyoxylate Shunt Enzymes as Persistence Targets: Structures of Isocitrate Lyase (Rv0867) and Malate synthase (Rv1837c) (Vivek Sharma and James C. Sacchettini)

Mycobacterium tuberculosis is a successful pathogen in part because it persists and maintains chronic infections in humans, notwithstanding an active immune response. *Mtb* persists by exhibiting diverse metabolic states, only few of which can be targeted by current antimycobacterials. While the growing bacteria can be eliminated by drugs specific for factors involved in cell growth and division, the slow-growing or dormant sub-populations maintain a sub-clinical infection weeks after the start of therapy. Therefore long (6-9 months) therapeutic regimens are necessary to ensure that all persisting bacteria enter a metabolic state that can be effectively killed by current drugs. Shortened therapeutic protocols will require new compounds that kill persistent bacteria.

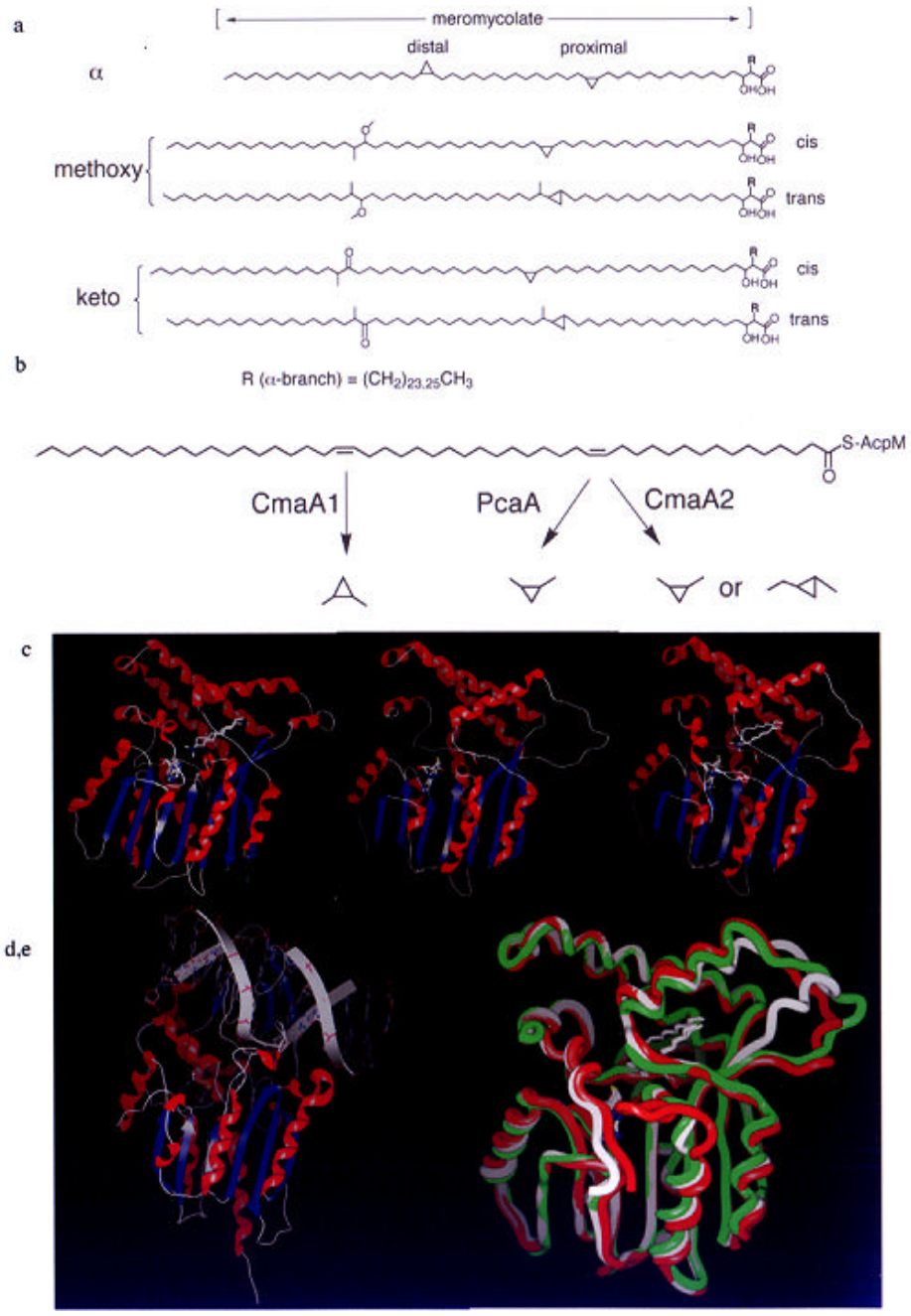

Fig. (4). cyclopropane synthases CmaA1, PcaA and CmaA2. (a), Structures of mycolic acids in *Mtb*. Alpha mycolates have a *cis* cyclopropanering at both the proximal and distal positions. Keto- and methoxymycolates have either a *cis* or *trans* cyclopropane ring at the proximal position and an oxygenated functional group at the distal position. (b)The reactions catalyzed by the cyclopropane synthases CmaA1, PcaA and CmaA2. These enzymes are thought to act on the meromycolate chain bound to acyl carrier protein (AcpM). (c) Ribbon representations of the structures of the mycolic acid cyclopropane synthases CmaA1, PcaA and CmaA2. α-helices are shown in red and β-strands in blue. CmaA1, shown in the first panel, was solved to 2. 0 Å resolution in complex with the product of methylation, *S*-adenosyl-L-homocysteine (SAH) and the lipid-like detergent didecyldimethylammonium bromide (DDDMAB). PcaA was solved in complex with SAH to 2. 0 Å resolution and is shown in the second panel. The structure of CmaA2 solved to 2. 65 Å resolution is shown in the third panel in complex with SAH and DDDMAB. (d) The structure of a ternary complex of HhaI methyltransferase with DNA and *S*-adenosyl-L-homocysteine. The cyclopropane synthases show a high level of structural similarity with other methyltransferases. Shown here is the structure of HhaI DNA methyltransferase for a comparison [64] , which like the cyclopropane synthases also has a core region containing a seven-stranded βsheet flanked by αhelices. (e) A superimposition of the three cyclopropane synthase structures. CmaA1 iscolored white, CmaA2 is red and PcaA is in green. The three structures are superimposed by the C_α atoms of the central α/βfold. The cofactor SAH and DDDMAB are shown in ball and stick representation. There is one region in which the structures of the proximal-acting enzymes differ from that of CmaA1, which can be seen in this superimposition and corresponds to residues 170-210 in PcaA.

Interestingly, persistence depends on genes required for supplying metabolites that allow mycobacteria to adapt to the adverse environment within the active macrophages. One strategy entails a metabolic shift in carbon source to C2 substrates generated by ,-oxidation of fatty acids [68]. Under these conditions, glycolysis is decreased and the glyoxylate shunt is significantly upregulated. The glyoxylate shunt converts isocitrate to succinate and glyoxylate by enzyme isocitrate lyase (ICL), followed by addition of acetyl-CoA to glyoxylate to form malate by malate synthase. The disruption of isocitrate lyase attenuates persistence of *Mtb* in mice or inflammatory macrophages [68]. iThe structures of both isocitrate lyase (Rv0867) and malate synthase (Rv1837c), the two enzymes of glyoxylate shunt, were solved using multi-wavelength anomalous dispersion (MAD) methods. ICL is a tetrameric protein (Fig. **5a**) with subunits of 428 amino acids. The striking features of this structure are inter-subunit helix-swapping and a large conformational change in the active site loop upon binding of substrates. The core of the structure is an unusual 8- (α/β) barrel where the eighth helix projectsaway from the barrel and forms interactions exclusively with the neighboring subunit [69]. As in most α/β barrels, the active site of ICL is located at the carboxy ends of the β-strands. The ICL signature sequence (K189KCGH193) containing the nucleophilic cysteine (cys191) is located on a flexible loop that undergoes a large conformational change (10-15Å) upon binding of substrates. This brings the cys191 next to the substrate and completely secludes the active site from bulk solvent.

The structures of ICL have also been done in complex with glyoxylate and with two inhibitors (3-nitropropionate and 3-bromopyruvate). These structures show how the substrates and products occupy the active site of ICL. While the glyoxylate binding deeper in the binding cavity by coordination to a Mg2+ ion, the succinate binds closer to the surface. In the bound form, the active site Cys191 is located 3-4 Å from the two atoms (C2 carbon of succinate and the aldehyde carbon of glyoxylate) undergoing the lysis of bond.

Malate synthase (malate synthase , Rv1837c) is the second enzyme of the glyoxylate pathway. Malate synthase catalyzes the Mg2+dependent condensation of glyoxylate and acetyl-coenzyme A and hydrolysis of the intermediate to yield malate and coenzyme A. The basic fold of malate synthase is also that of an 8- (·/,) barrel. The centrally located barrel is flanked by two ·-helical domains and contains an insertion of a ,-sheet domain on one side of the barrel (Fig. **5b**). Several inhibitors have been identified against malate synthase and ICL using structure-based design and high-throughput screens and studies to characterizethe structures of these inhibitor complexes are underway. Such inhibitors have potential to be developed into drugs against persistent bacteria.

4. Structure Determination of *Mycobacterium tuberculosis* InhA:Inhibitor Complexes (Mack Kuo and James C. Sacchettini)

Fatty acid biosynthesis is responsible for the production of the precursors of mycolic acids, and therefore represents an attractive drug target. To date, the most effective drugs against *M. tuberculosis* inhibit proteins responsible for mycolic acid and cell wall biosynthesis [70, 71]. InhA, a mycobacterial enoyl-acyl carrier protein (ACP) reductase, occupies a prominent role in FAS-II fatty acid biosynthesis. InhA is the target of the first-line antitubercular drug

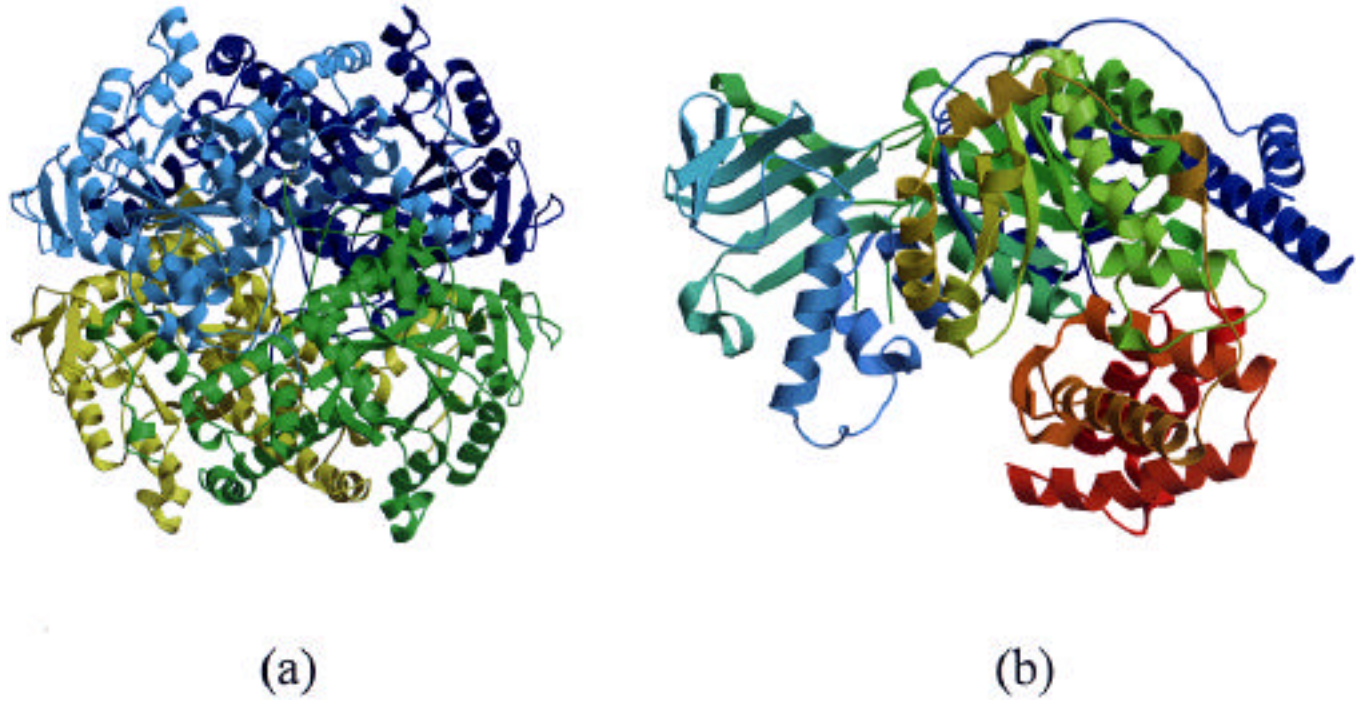

Fig. (5). Three-dimensional structures of glyoxylate pathway enzymes. (a) Ribbon representation of the tetramer of isocitrate lyase (ICL). Each subunit is shown in different color. (b) Ribbon representation of malate synthase.

isoniazid (isonicotinic acid hydrazide; INH). Inactivation of InhA alone is sufficient to inhibit mycolic acid biosynthesis and induce cell lysis [72]. As the FAS-II system is absent in eukaryotes, but is essential to the viability of *Mtb*, components of the FAS-II system such as InhA represent superb drug targets.

InhA has an amino acid sequence and structure similar to those of two previously characterized enoyl-ACP reductases: *Escherichia coli* FabI [73] and *Brassica napus* ENR [74]. The antibiotic triclosan (2,4,4'-trichloro-2' hydroxydiphenyl ether) specifically targets and inhibits the enoyl-ACP reductase component of *E. coli* [75, 76]. The structures of the *E. coli* FabI:triclosan complex [77, 78, 79] and *B. napus* ENR:triclosan complex [80], and the observation that triclosan inhibits *M. tuberculosis* InhA [81] led us to initiate structural studies on the *M. tuberculosis* InhA:triclosan complex in order to elucidate the molecular basis of triclosan inhibition.

The structure of the ternary complex of InhA:NAD:triclosan revealed a molecule of triclosan situated in the active site in the same orientation as in the *E. coli* and *B. napus* enoyl-ACP reductases (Fig. **6a**). Triclosan is positioned adjacent to the NAD cofactor, with the 2-hydroxyphenyl ring stacked on the nicotinamide of the nucleotide cofactor. The 2 hydroxyl group of triclosan also forms hydrogen bonds with Tyr158 and the 2'-hydroxyl group of the nicotinamide ribose. However, in contrast to the *E. coli* FabI and *B. napus* ENR, a second molecule of triclosan was present, inverted above the first, which stabilized the fatty-acyl substrate binding loop. A multiple sequence alignment of enoyl reductase homologues revealed that mycobacteria inhA contains an additional 10 residues in their substrate binding loop, consistent with their role in synthesizing long chain fatty acids. Superposition of the previously determined InhA:NAD:C16 fatty acyl substrate ternary complex and the InhA:NAD:triclosan ternary complex disclosed that the C16 fatty acyl substrate and the two molecules of triclosan occupy a nearly identical orientation within the active site.

We have also determined the co-crystal structures of *Mtb* InhA with NAD and with a compound (Genz-10850) identified through high throughput screens. In the co-crystal structure, Genz-10850 lies above the NAD cofactor, and its carbonyl group forms hydrogen bonds with Tyr158 and the 2'-hydroxyl oxygen of the nicotinamide ribose of NAD. The NH of the indole ring of the drug interacts with a phosphate oxygen of NADH (Fig. **6b**). The highly hydrophobic fluorenyl group is shielded from solvent by the fatty-acyl substrate binding loop of InhA.

A promising strategy for structure-based drug design is to construct compounds that stack with Phe149 (mimicking the isoniazid mechanism), that hydrogen-bond with Tyr158 (like triclosan), that interact with the NADH cofactor (through stacking and/or hydrogen bonding), and that stabilize the fatty acyl substrate binding loop (in the same manner as Genz-10850, the fatty acyl substrate, and triclosan). Triclosan analogues, tethered triclosan dimers, and high throughput screen lead compounds could be used as a molecular scaffold for further studies. Inhibitors based on these premises may have several advantages: they will not require activation (unlike isoniazid), and they may specifically target mycobacteria, as enoyl reductases from other organisms have much shorter fatty-acyl binding loops.

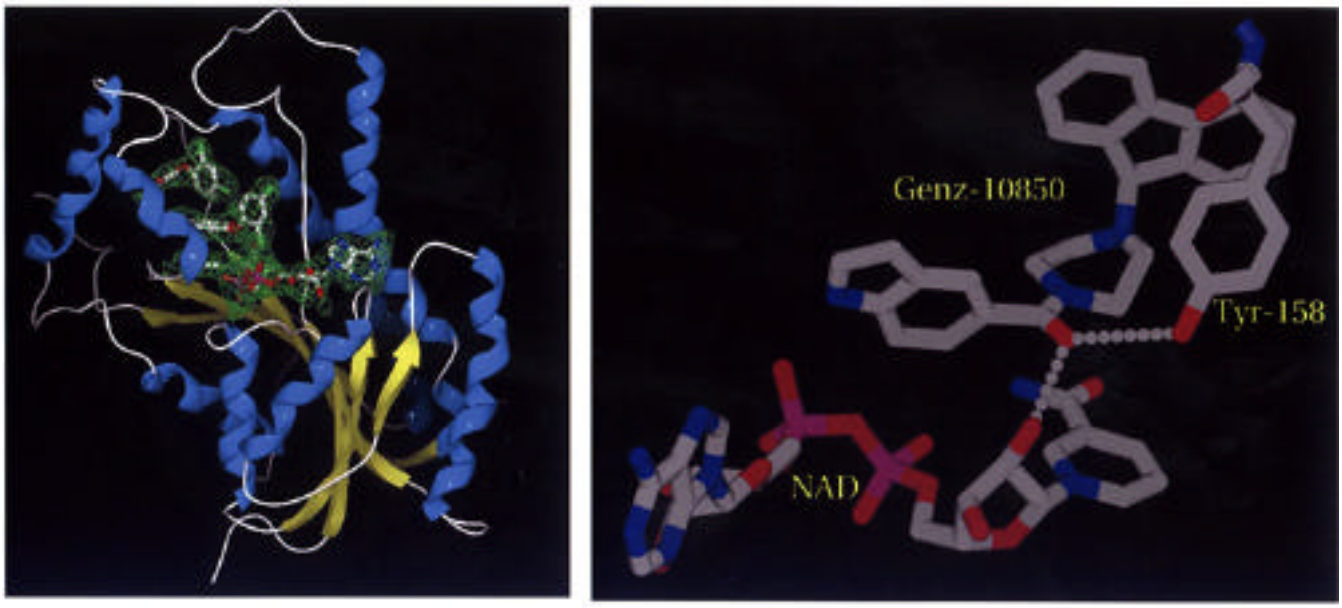

Fig. (6). Three-dimensional structures of InhA:inhibitor complexes. (a) Ribbon representation of a monomer of InhA bound to two molecules triclosan. (b) Active site of the InhA:Genz-10850 complex.

5. Structural Studies on the Rv2002 Soluble Mutant (Jin Kuk Yang, Min S. Park,Geoffrey S. Waldo & Se Won Suh)

The Rv2002 gene (fabG3) from *Mycobacterium tuberculosis* encodes a 260-residue protein, with a calculated molecular mass of 27 030 Da. It contains the dinucleotide binding motif GXXXGXG (14-20) and YXXXK (153-157) sequence motif, which are two highly conserved features of the short-chain dehydrogenase/reductase (SDR) family enzymes. On the basis of sequence identity (31%), it has been annotated as a homolog of *M. tuberculosis* -ketoacyl carrier protein reductase, which is the second enzyme in fatty acid elongation cycle and is encoded by the fabG1 gene (Rv1483) [82]). Among -ketoacyl carrier protein reductases, the highest sequence identity is observed with that from *Bacillus halodurans* (TrEMBL, Q9KA03) (38% in a 244-residue region of overlap). Rv2002 also shows significant sequence similarity with other enzymes such as 3, 20-hydroxysteroid dehydrogenase (49% identity in a 243-residue overlap with the one from *Streptomyces hydrogenans*; Swiss-Prot P19992) and l-3-hydroxyacyl-CoA dehydrogenase of the fatty acid -oxidation pathway (35% identity in a 204-residue overlap with the one from rat brain; Swiss-Prot, O70351). Therefore, its molecular or biological function cannot be unambiguously assigned on the basis of sequence similarity alone.

The Rv2002 gene product was expressed in *Escherichia coli*, but found to be in an insoluble form within inclusion bodies. Such insolubility is a common problem when expressing heterologous proteins. In some cases site-directed mutations are introduced to improve solubility, but this procedure involves extensive trial-and-error. To create a rapid, general method to find mutants that improve expression and solubility, a green fluorescent protein (GFP) based method has been developed [83]. The essence of the method is that if GFP is expressed as a C-terminal fusion attached to the protein of interest, the fluorescence of GFP is a reporter of the folding of both proteins. This can be used to engineer mutants of the protein of interest that fold robustly even in the absence of the GFP reporter.

We applied the GFP-based directed evolution technique in order to obtain soluble mutants of Rv2002. After three rounds of forward evolution, several soluble mutants with three to five point mutations were obtained. One of the soluble mutants with three point mutations, I6T/V47M/T69K, was overexpressed for structural studies. All three mutations are outside the conserved regions of the SDR family and were therefore expected to have minor effects on the catalytic activity. Crystals of the soluble mutant, diffracting to 1.8 Å resolution,were obtained with precipitation using polyethylene glycol 3000. The crystals belong to the space group *P*3221, with cell dimensions of $a = b = 70.38$ Å and $c = 148.93$ Å [84]. The structure was solved using the Se MAD method in the NAD+-bound state. In the crystal, four identical subunits form a tetramer with the 222 symmetry. Each subunit comprises a single domain, a characteristic dinucleotide binding fold (Rossmann fold), of dimensions 40 Å x 45 Å x 50 Å. The crystal structure reveals that Asp38 plays a key role in determining the cofactor preference of NADH over NADPH, which has been confirmed by enzymatic assays. The cofactor preference suggests that the Rv2002 gene product is unlikely to function *in vivo* as -ketoacyl carrier protein reductase. Further experiments are necessary to establish its physiological role.

6. Exploring New Drug targets of *Mycobacterium tuberculosis* (Radha Chauhan, Avinash Kale, Nandita Bachhawat & Shekhar C. Mande)

Inositol-1-Phosphate Synthase

Inositol and its metabolites are widely found in eukaryotes but are uncommon in prokaryotes. As a consequence, prokaryotic inositol metabolism remains poorly understood. The first known enzymatic step in the *de novo* biosynthesis of inositol is mediated by the *ino1*gene, which encodes 1L-*myo*-inositol-1-phosphate synthase (I1-P synthase, [85]. In the presence of NAD (nicotinamide adenine dinucleotide), this enzyme catalyses the conversion of glucose-6-phosphate (G-6-P) into inositol-1-phosphate. The inositol-1-phosphate is subsequently dephosphorylated to *myo*-inositol by a specific I1-P phosphatase [86, 87]. In yeast and fungi, genetic evidence indicates that I1-P synthase is essential for viability [88].

In mycobacteria, several important metabolites such as phosphatidylinositol (PI), phosphatidylinositol mannosides (PIM), and glycosylphosphatidylinositol (GPI) are derived from inositol. Thus, *de-novo* biosynthesis of inositol might be a useful pathway to intercept for drug development. We recently identified a I1-P synthase in *Mycobacterium tuberculosis H37Rv* (Rv0046c) through a combination of computational analysis and complementation in yeast, leading to the first evidence for the presence of this enzyme in a prokaryote [89, 90].

To look for genes that might be involved in inositol metabolism in *Mtb*, the mycobacterial genome database was searched with the BLAST program [91] using the *S. cerevisiae* inositol-1-phosphate synthase (INO1p) amino acid sequence. The search revealed a very low sequence homology (less than 15%) to the Rv0046c open reading frame of the *M. tuberculosis H37Rv* genome coding for a 367 amino acid long protein. However the absence of the sulfhydryl groups known to be important for the enzymatic activity of eukaryotic INO1p, the lack of the GXGXXG NAD-binding motif, and the failure of a PROSITE search to identify nucleotide binding motifs [92] argued against the identification of Rv0046c as inositol-1-phosphate synthase. Nonetheless, a threading calculation predicted a Rossmann fold in the N-terminal half of the Rv0046c protein sequence, suggesting that this domain might be functionally important for NAD or NADP binding [89].

A yeast based functional assay system was employed to test whether Rv0046c encodes an inositol-1-phosphate synthase enzyme. A yeast strain lacking the gene for yeast *INO1*showed that Rv0046c is capable of functionally complementing the yeast *ino1* mutation [93, 89]. The computer analysis and complementation assay in *S. cerevisiae* therefore indicate that Rv0046c ORF indeed has the INO1-P synthase activity.

Hydroperoxide Reductase

Isoniazid has been one of the frontline drugs used in tuberculosis control for more than three decades. Resistance to isoniazid develops primarily due to point mutations in the *katG*locus, encoding the catalase peroxidase enzyme [94]. The loss of *katG* function in isoniazid strains therefore renders *Mtb* susceptible to the toxic peroxide radicals. Under these conditions, overexpression of another gene, *ahpC*, encoding an alkyl hydroperoxide reductase, has been frequently found to occur and might help bacteria to overcome stress [95]. Consistent with this, antisense RNA studies have shown that mycobacterial AhpC is important for virulence [96].

The mycobacterial AhpC differs in many ways from its well-characterized homologues in enteric bacteria. For example, the *Mtb* enzyme possesses three cysteine residues in its active site, compared to two in the *Salmonella typhimurium* and *Escherichia coli* enzymes [97, 98]. An additional difference between the *Mtb* AhpC and that derived from other closely related species is that the electron donor partner of AhpC, the FAD/NADPH binding subunit AhpF, is absent in *Mtb*.

We have generated active site cysteine to alanine mutants of AhpC (Rv2428) to probe their role in activity. Four mutants (C61A, C174A, C176A and C174/176A) were studied. The quaternary state of wild type and each mutant was characterized by gel filtration chromatography and SDS-PAGE. Their comparison is summarized in Table **1**. The *Mtb* AhpC is a decameric enzyme composed of five identical dimers. The dimer formation is mainly due to intersubunit disulfide linkage, and the further oligomeric interactions are dominated by ionic character. During oxidation and reduction of enzymatic cycle, AhpC undergoes conformational changes. We found that cysteine to alanine mutations have significant effect on oligomerization of AhpC [99].

The four mutants were also characterized for biochemical activity (Table **1**). The table also lists the number of free sulfhydryl groups in each of the mutants, and their oligomeric state. The assays showed that mycobacterial AhpC has general antioxidant properties [100]. This property might be a favorable characteristic for mycobacterial survival within macrophages in the absence of *katG* in isoniazid resistant strains. The kinetic parameters of mutants suggest that all three cysteine take part in enzyme activity [100].

On the basis of our site directed mutagenesis experiments we were able to propose a mechanism of action for mycobacterial AhpC which appears to be different from 1-Cys and 2-Cys Prx mechanisms (Fig. **7**). A unique disulfide relay mechanism was observed in mycobactrial AhpC suggesting that it might belong to a distinct class of peroxiredoxins. Our phylogentic studies suggested that this enzyme might have evolved from a eukaryotic 2-Cys Prx.

To gain further structural insights into the active site of mycobacterial AhpC, a homology model for *Mtb* AhpC was constructed. We used two 2-Cys Prx three-dimensional structures as templates for homology modeling, thioredoxin peroxidase B from red blood cells (pdb code 1QMV) and mammalian 2-Cys peroxiredoxin, Hbp23 (pdb code 1QQ2) [101, 102]. The thioredoxin peroxidase is a decameric structure and the peroxiredoxin is a dimer. The active site of alkylhydroxyperoxidase reductase (AhpC) is essentially a hydrophobic pocket, but residues adjoining the active site cleft include buried charged amino acids. The three cysteine residues (Cys 61, 174 and 176) are near the cleft with Cys 174 buried, Cys 176 partially exposed, and Cys 61 residue exposed in the cavity. Either of two of the cysteines (174 and 176) could possibly form an intersubunit disulfide bond with Cys 61 of another subunit. Overall our model indicates that the mycobacterial AhpC has unique active site which may be different than eukaryotic peroxiredoxins.

7. Rv3853 (MenG): A Putative Drug Target that Appears to be a case of Mistaken Functional Identity. (Jodie M. Johnston, J. Shaun Lott, Edward N. Baker and Vickery L. Arcus)

Menaquinone (Vitamin K) is an essential vitamin that is an obligatory component of the anaerobic electron transfer pathways that operate not only in strict anaerobes, but also in aerobic gram-positive bacteria including *Mtb* [103]. This function may be particularly important in *Mtb* under conditions of low oxygen, and may thus play a role in the

Table 1. Activities of AhpC Mutants

Name of Protein	Oligomerization	* Relative Activity	Free -SH per Enzyme Molecule**
AhpC: wild type	Decamer	100	0.77+0.2
AhpC:C61A	Dimer	0	1.5+0.28
AhpC:C174A	Dimer	0-5	0.14+0.17
AhpC:C176A	mixed	40-65	0.14+0.19
AhpC:C174/176A	dimer	20-40	0.825+0.183

* represents average of the activity determined by four assays [100].

** free -SH groups were determined by the DTNB method [100].

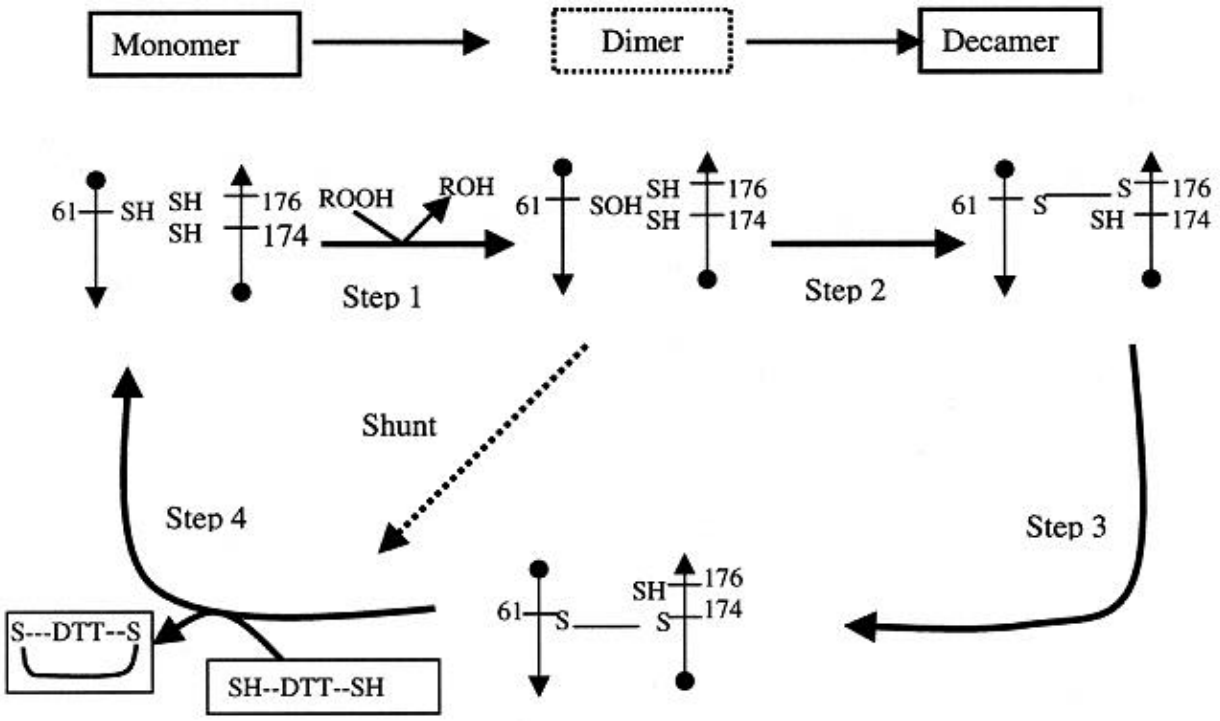

Fig. (7). Proposed Reaction mechanism of mycobacterial AhpC: Step1, Attack on Cys 61 by peroxide substrate (ROOH) to form oxidized intermediate and alcohol (ROH) as products; Step 2, formation of intersubunit disulfide bond between C61 and C176; Step 3, formation of C61-C174 intersubunit disulfide bond; Step 4, reduction of intersubunit disulfide bond with DTT. Shunt path is a direct reduction of the oxidized Cys61 intermediate by DTT which is observed in C174/176A mutant. The fully reduced enzyme might be dimeric, while oxidized enzyme might be decameric in nature.

phenomenon of persistence, when the bacteria reside in activated macrophages. Coupled with the observation that menaquinone is an essential nutrient that is not synthesized in animals, this makes enzymes of the menaquinone biosynthetic pathway attractive drug targets.

In *E. coli* the menaquinone biosynthetic pathway involves either 7 or 8 enzymes [103]. Sequence comparisons have found homologues for 7 of these enzymes in the *Mycobacterium tuberculosis* H37Rv strain [14]. One of these proteins, Rv3853, has been annotated as MenG, based on sequence similarity to an *E. coli* enzyme that was thought to be the terminal enzyme in the pathway. MenG is proposed to be the SAM-dependent methyltransferase that transfers a methyl group to demethyl menaquinone in this last step [104]. Intriguingly, the MenG sequence has none of the common methyltranferase motifs, and the *Mtb* genome encodes another protein, identified as UbiE, that could also catalyse this final step [105, 106]. To clarify its function by revealing possible homologies not seen at the primary sequence level and to provide a template for possible drug design, determination of the structure of the gene product of Rv3853 was undertaken in the context of the *Mtb* structural genomics effort.

The Rv3853 gene from *M. tuberculosis* was amplified from *Mtb* genomic DNA and expressed in *E. coli* as a His-tagged protein. Expression tests showed the protein to be insoluble under all conditions tried, so we purified it from inclusion bodies and then refolded it *in vitro.* The refolded protein was purified and then was crystallized from a solution containing 0. 45 M Na/K tartrate. The MenG structure was determined by SIRAS (single isomorphous replacement with anomalous scattering) from an Hg derivative, and refined at 1.90 Å resolution.

The MenG monomer was found to have a fold that has been seen in other proteins. The fold is described in SCOP [107] as a 'swivelling 'β/β/α fold, and in CATH [108] as a 3-layer β/β/α sandwich. In this structure, the first layer consists of a 4-stranded antiparallel β-sheet (S1, S11, S10, S2, see Fig. **8**) sitting adjacent to a 2-stranded antiparallel β-ribbon (S8, S9). This layer packs against a central 5-stranded mixed β-sheet (S3, S4, S5, S6, S7) that forms the second layer. The third layer of the "sandwich" comprises three parallel helices that provide the S3-S4, S4-S5 and S5-S6 connections. A large loop (B) and short strand (S7) wrap around layers 2 and 3, and connect to layer 1. Three MenG monomers further associate in "nose-to-tail" fashion into a doughnut-shaped trimer with a large hole in the center (see Fig. **9**).

Surprisingly, the Rv3853 structure suggested that it is not a methyltransferase. The SAM-dependent methyltransferases have a characteristic fold, with a mixed 7-stranded β-sheet flanked by helices and a conserved S-adenosylmethionine binding pocket at the C-terminal end of the β-sheet [65]. This conserved binding pocket provides a signature sequence by which new methyltransferases may be identified. Indeed, the structure of a conserved hypothetical protein from *M. tuberculosis* (Rv2118c) was recently determined and identified as a tRNA methyltransferase based

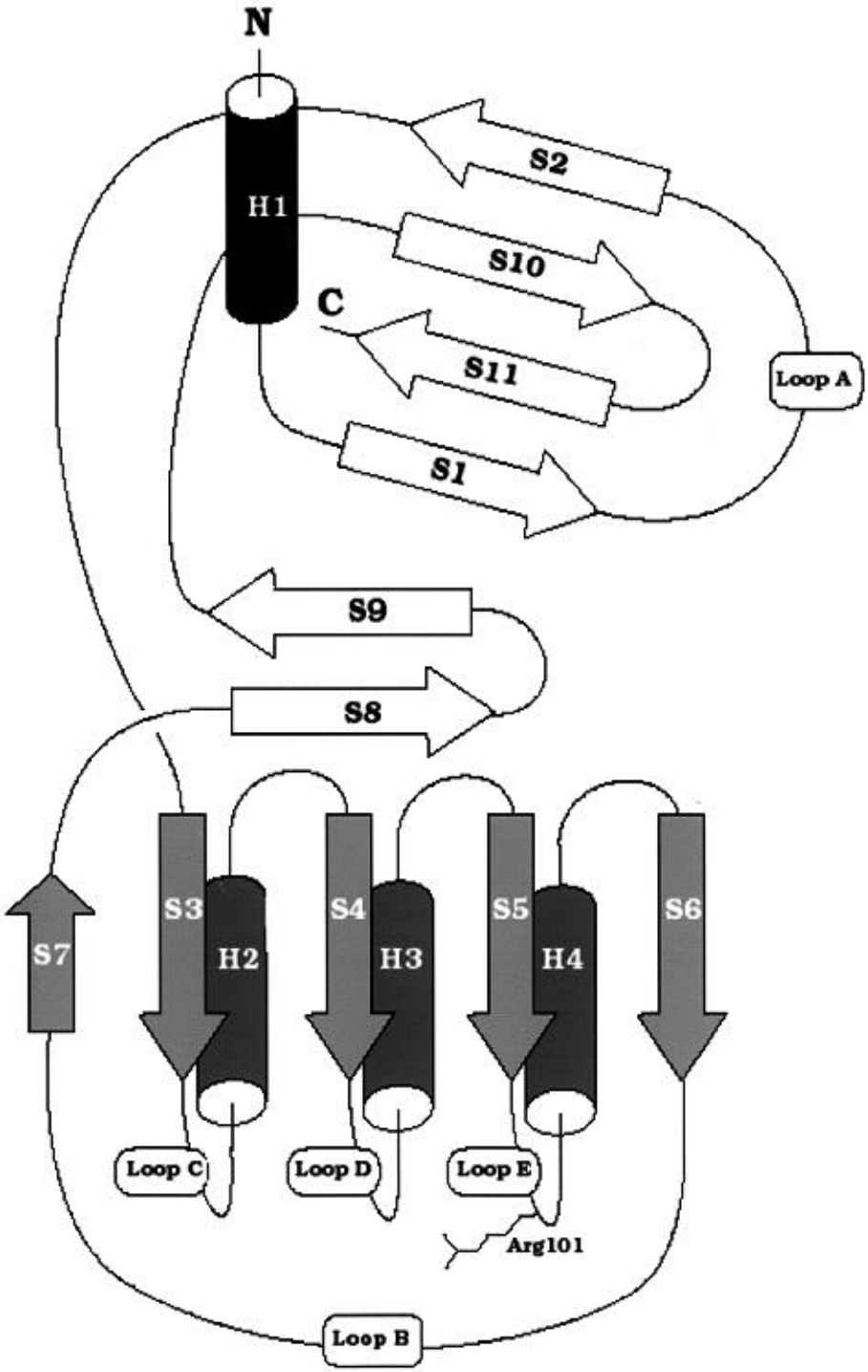

Fig. (8). Topology diagram of Rv3853. The layers referred to in the text are differentiated by shading: Layer 1 = white; Layer 2 = light grey; Layer 3 = dark grey.

on the methyltransferase signature sequence and structural conservation with this family [109]. The structure of Rv3853 shows neither the charcteristic fold nor the conserved binding pocket.

What can be learnt from the Rv3853 structure? Comparisons of the monomer with all protein structures in the Protein Data Bank, using DALI ([110], revealed five domains with significant structural similarity. The closest match is to the phosphohistidine transfer domain of pyruvate phosphate dikinase [111]. This has a characterized fold, and protein structures classified under this fold in SCOP include carbamoyl phosphate synthetase and an aconitase domain [107, 111]. The structural similarity between the phosphohistidine transfer domain from pyruvate phosphate dikinase and Rv3853 is sufficiently low that it is unlikely that Rv3853 is a member of this superfamily. The topology is well conserved but there is no sequence conservation at the phosphohistidine active site or elsewhere. The structural homology does, however, point to a likely location for the

Fig. (9). Schematic ribbon diagram of the Rv3853 trimer. The view is down the crystallographic three-fold axis. The diagram was drawn using Molscript and Raster3D.

active site of Rv3853. The position of the active site histidine in the phosphohistidine transfer domains coincides closely with the location of an arginine residue in Rv3853 (Arg101) which is strictly conserved amongst homologues of Rv3853. In addition, this arginine residue appears to hydrogen bond to a ligand, possibly tartrate from the crystallization medium, that is bound nearby. These two pieces of evidence, combined with the presence of a negatively charged canyon adjacent to Arg101, lead us to the hypothesis that this forms the active site. The substrate and enzymatic activity remain unknown, however.

The structure of Rv3853 illustrates what is likely to be a common problem in choosing targets for structure-based drug design. Rv3853 and its homologues have all been annotated as MenG, implying that they catalyse the last step in menaquinone biosythesis. The crystal structure of Rv3853 suggests that this may be a case of an incorrect annotation propagating through a series of databases. The methyltransferase signature sequence is absent and the structure belongs to a different fold whose closest structural homologues are phosphohistidine transfer domains. In addition, we have soaked our crystals in both S-adenosyl methionine and an isoprenyl mimic and neither appears to

bind to our protein in the crystal. We also conclude that the other *Mtb* enzyme previously annotated as possibly catalyzing the last step in menaquinone biosynthesis (Rv0558, UbiE) is the likely candidate for this enzymatic activity.

8. Structural Analysis of *M. tuberculosis* Cytochrome P450 Enzymes (David Leys, Kirsty J. McLean and Andrew W. Munro)

The cytochromes P450 (P450s) are members of an enzyme superfamily found in all forms of life. The P450s have a cysteinate-ligated *b*-type heme. They catalyse the reductive scission of molecular oxygen, usually with resultant mono-oxygenation of a hydrophobic substrate (steroids, polyketides and fatty acids are typical substrate classes) and production of a molecule of water. The general scheme of the reaction is

$$R + O_2 + 2e^- + 2\,H^+ \rightarrow ROH + H_2O.$$

where R = the substrate and ROH = the hydroxylated product.

The reaction requires two protons (delivered from the solvent) and two electrons (e-), which are delivered ultimately from NAD (P)H via flavoprotein or flavoprotein/ferredoxin redox partners.

P450 enzymes are widespread in eukaryotes and perform essential roles in human steroid synthesis and drug metabolism. They are much rarer in prokaryotes, where they participate in pathways for the degradation of unusual organic compounds such as camphor [112]. More than 20 P450s were predicted from the genome sequence of *Mtb* [14], implying that there are essential roles for P450 enzymes in the physiology of the pathogen. Subsequently, the genome sequence of another actinomycete (*Streptomyces coelicolor*) has indicated the presence of many P450 enzymes, suggesting that this genus may have P450-dependent metabolism very different from other bacterial genera.

The structures of two of the *Mtb* P450s have now been solved. The first is CYP51 [113], the product of the Rv0764c gene (PDB codes 1E9X, 1EA1). CYP51 is homologous to eukaryotic sterol demethylases including obtusifoliol demethylases in plants and lanosterol demethylases in fungi. The *Mtb* CYP51 protein is a 51 kDa enzyme which catalyses 14α-demethylation of lanosterol and obtusifoliol. Its activity can be examined *in vitro* using *E. coli*NADPH flavodoxin reductase and flavodoxin as redox partners [114]. The catalytic turnover of the enzyme in this system is low, but the fact that both sterols could be demethylated suggests that sterol metabolism may be a feature of *Mtb*. The presence of sterols in *Mtb* remains to be proven.

The second P450, CYP121, is the product of the Rv2276 gene (D. Leys *et al.* , paper in preparation). This 43 kDa enzyme has an amino acid sequence similar to prokaryotic P450s involved in polyketide antibiotic synthesis, including the structurally characterized P450 eryF (CYP107A1) that catalyzes hydroxylation of 6-deoxyerythronolide B in the final stage of erythromycin biosynthesis in *Saccharopolyspora erythraea* [115, 116]. Thus it appears likely that CYP121 has a role in polyketide metabolism, or possibly oxygenation of an alternative polycyclic molecule.

The Mtb CYP51 structure has been solved in the presence of heme-ligating azole drugs (4-phenylimidazole and fluconazole), while the CYP121 structure is in a ligand-free form.

Only a few P450 structures have been solved (seven are in the public domain), and the *Mtb* P450s are the first two solved from the same organism. Moreover, both of the *Mtb* P450s are potential drug targets. Azole anti-fungal drugs (e. g. fluconazole, ketoconazole, clotrimazole) were designed as inhibitors of the fungal CYP51 sterol demethylase enzymes, which catalyze the production of the essential membrane ergosterol from the substrate lanosterol [117]. Since it was thought that such CYP51 P450s (and sterol pathway) were absent from bacteria, no systematic investigation of such azole drugs as anti-mycobacterial agents has been carried out. However, it is now known that drugs in this class bind tightly to the *Mtb* proteins CYP51 and CYP121, leading to inactivation of the P450s. The fact that both these *Mtb* P450s have high affinity for azole drugs reinforces the potential of this type of antibiotic approach against the pathogen.

The two *Mtb* P450s CYP51 and CYP121 are similar in overall topology both to each other and to other known P450 structures (Fig. **10**). Both are predominantly alpha helical and their "I" helices contain a motif of several amino acids central to substrate selectivity and proton transfer in all known P450s [118]. In CYP51 the N-terminal portion of this helix is kinked more substantially than in CYP121 (or any other P450) which may provide an increased binding space for inhibitors. In one of the inhibitor-bound structures fluconazole is bound in the active site cavity such that the triazole ring is positioned perpendicularto the porphyrin plane with a ring nitrogen atom coordinated tothe heme iron, and with F83 and F255 providing additional non-bonding contacts. The main fluconazole-induced conformational changes,in comparison to the smaller4-phenyl imidazole ligand in the other azole-bound structure, involve a helix-coil transition of the C helix and displacement of the residues in the I helix.

In CYP121, the water-filled active site pocket is one of the largest of any P450 (approx 1350 Å3). Key players in substrate interactions appear to be Arg386 and Ser237, which approach close to the distal face of the heme and are hydrogen bonded to one another in the absence of substrate. Such a motif has the potential to bind to a substrate carboxylate group. The high resolution obtained for this structure (1.2 Å) has allowed identification of a bifurcated proton delivery pathway, which might well function to deliver the two protons needed by two separate, converging routes. Furthermore, acetate and iodopyrazole (used as derivative in the MIRAS structure determination) bind in a narrow funnel-shape region of the active-site cavity between

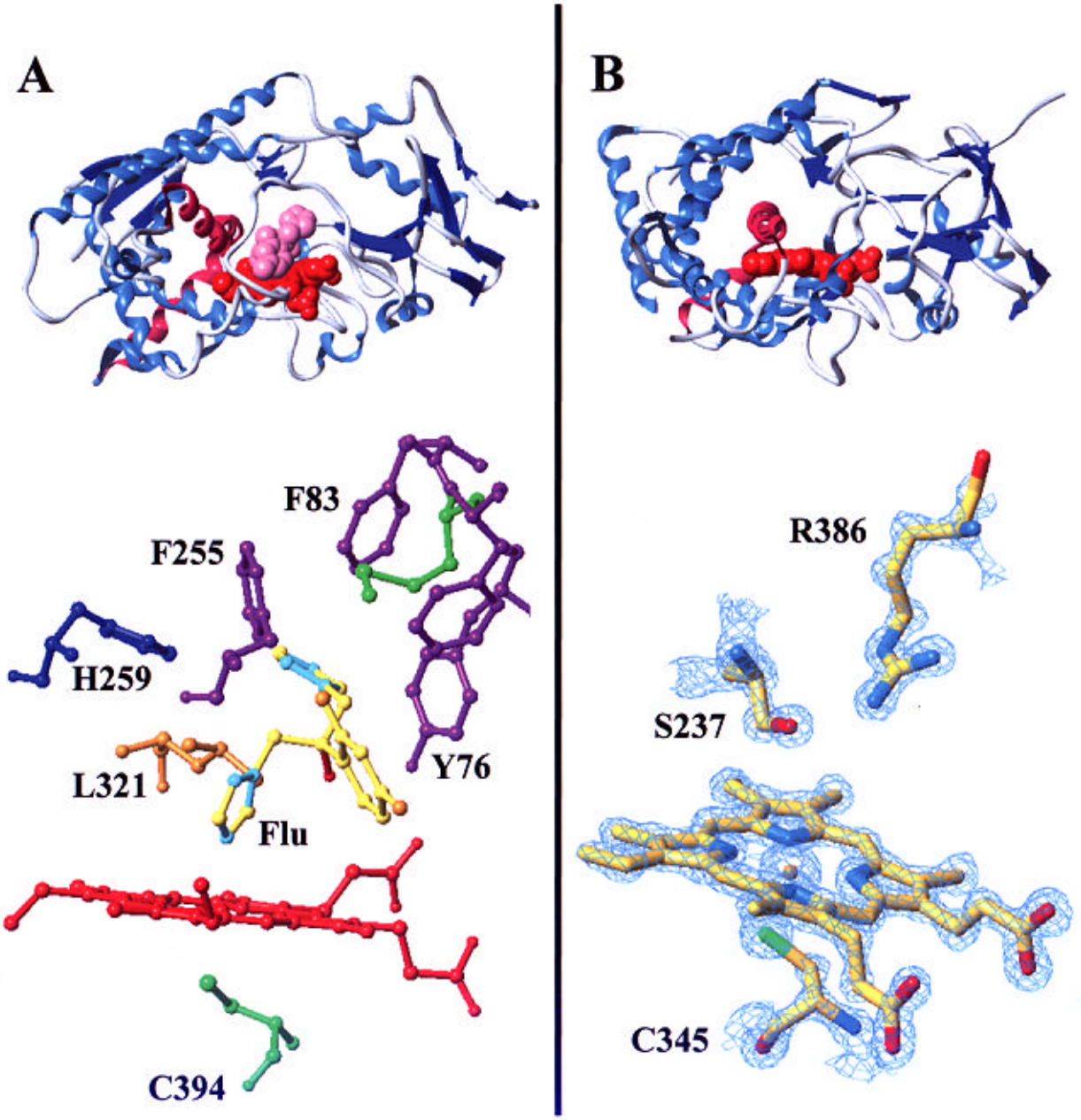

Fig. (10). Comparison of two *Mtb* P450 enzymes. The overall topology (above) and active site (below) of the two structurally characterised *Mtb* P450s: the sterol demethylase (CYP51) (A) and CYP121 (B). The folds are similar in the two P450s (heme in red spacefill, secondary structural elements in blue, fluconazole inhibitor in lilac), although the major helical feature (the I helix, shown with its adjacent J helix in magenta) is severely "kinked" on the N-terminal side for CYP51. Residues involved in CYP51 binding to the fluconazole are depicted, coloured according to residue type. The imidazole N3 of the inhibitor hydrogen bonds with H259, and several non-bonding contacts are made with other side chains. For CYP121, residues R386 and S237 inferred in binding the substrate are shown, along with the heme iron cysteine ligand (C354), surrounded by the 1. 2 Å 2Fo-Fc map contoured at 2 sigma. The heme iron ligand in CYP51 is C394. The sterol demethylase active site is smaller and more hydrophobic than is the CYP121 active site. This indicates that the substrate for CYP121 may be bulky and at least partly of hydrophilic character, possibly a polyketide.

Phe168 and Trp182, suggesting a possible recognition site for the substrate. At present we are pursuing the structure of CYP121 in complex with inhibitors and putative substrates, a goal that will provide us with the necessary data to initiate structure-based drug-design. The high reproducibility and resolution attained for the CYP121 crystals will no doubt assist in the timely elucidation of the CYP121-inhibitor complex structures.

Perspectives

The TB Structural Genomics Consortium is applying a methodical approach to developing anti-TB drugs through the accumulation of functional and structural information of *M. tuberculosis* proteins. This review has presented the concerted efforts of a team of biologists and biochemists around the world in elucidating and functionally

characterizing protein structures essential to the growth and virulence of *M. tuberculosis*. The project is still in its infancy, but demonstrates the potential of structural genomics in providing a foundation for rational drug discovery. Structural and functional information determined through this structural genomics effort will enable use of this information to develop new specific drugs and drug treatments for tuberculosis.

ACKNOWLEDGEMENTS

Some of the authors wish to acknowledge Mari Gingery, Cameron Mura, Duilio Cascio, Marila Gennaro and Michael Sawaya for useful suggestions and discussion. The coordinates for Rv1886c have been deposited in the Brookhaven Protein Data Bank with the accession number 1FOP. Additional PDB accession numbers are 1KP9, 1KPG and 1KPH for CmaA1, 1KPI for CmaA2 and 1L1E for PcaA. Parts of this work have been supported by grants from the Department of Energy and the National Institutes of Health, by the New Economy Research Fund of New Zealand, and by the Korea Ministry of Science and Technology (NRL-2001, grant no. M1-0104000132-01J000006210).

ABBREVIATIONS

ACP = Acyl carrier protein

ATP = Adenosine triphosphate

CFA synthase = Cyclopropane fatty acid synthase

CTAB = Cetyltrimethylammonium bromide

DDD MAB = Didecyldimethylammonium bromide

G-6-P = Glucose-6-phosphate

GFP = Green fluorescent protein

GPI = Glycosylphosphatidylinositol

ICL = Isocitrate lyase

INH = Isonicotinicacidhydrazide

INO1p = Inositol-1-phosphate

MAD = Multiwavelength anomalous diffraction

MPD = Methylpentanediol

PI = Phosphatidylinositol

PIM = Phosphatidylinositol mannosides

= S-adenosyl-L-methionine

SAH = S-adenosyl-L-homocysteine

SAM-Mtases = S-adenosyl-L-methionine methyltransferases

SDR = Short-chain dehydrogenase/reductase

TB = Tuberculosis

Mtb = Mycobacterium tuberculosis

REFERENCES

[1] Terwilliger, T. C.**;** Waldo, G.**;** Peat, T. S.*;* Newman, J. M.*;* Chu, K.; Berendzen, J. *Protein Science*, **1998**, *7*, 1851.

[2] Sali, A. *Nature Struct. Biol*, **1998**, *5*, 1029.

[3] Burley, S. K. *Nature Struct. Biol*, **2000**, *7*, 932.

[4] Pellegrini, M.*;* Marcotte, E. M.*;* Thompson, M. J.*;* Eisenberg, D.; Yeates, T. O. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 4285.

[5] Marcotte, E. M.*;* Pellegrini, M.*;* Ng, H. L.*;* Rice, D. W.*;* Yeates, T. O.; Eisenberg, D.*; Science*, **1999**, *285*, 751.

[6] Dandekar, T.*;* Snel, B.*;* Huynen, M.; Bork, P. *Trends Biochem. Sci*, **1998**, *23*, 324.

[7] Overbeek, R.*;* Fonstein, M.*;* D'Souza, M.*;* Pusch, G. D.; Maltsev, N. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 2896.

[8] Eisenberg, D.*;* Marcotte, E. M.*;* Xenarios, I.; Yeates, T. O. *Nature*, **2000**, *405*, 823.

[9] Marcotte, E. M.*;* Pellegrini, M.*;* Thompson, M. J.*;* Yeates, T. O.; Eisenberg, D. *Nature*, **1999**, *402*, 83.

[10] Guilhot, C.*;* Jackson, M.; Gicquel, B. In *Mycobacteria: Molecular Biology*; *Virulence*, **1999**, Blackwell Science Ltd., 17.

[11] Bruccoleri, R. E.*;* Dougherty, T. J.; Davison, D. B. *Nucleic Acids Res.* , **1998**, *26*, 4482.

[12] Thole, J.*;* Janssen, R.; Young, D. *Mycobacteria: Molecular Biology, Virulence*, **1999**, Blackwell Science Ltd., 356.

[13] Webb, V.; Davies, J. in *Mycobacteria: Molecular Biology*; *Virulence*, **1999**, Blackwell Science Ltd., 287.

[14] Cole, S. T.*;* Brosch, R.*;* Parkhill, J.*;* Garnier, T.*;* Churcher, C.*;* Harris, D.*;* Gordon, S. V.*;* Eiglmeier, K.*;* Gas, S.*;* Barry, C. E.*;* 3rd, Tekaia, F.*;* Badcock, K.*;* Basham, D.*;* Brown, D.*;* Chillingworth, T.*;* Connor, R.*;* Davies, R.*;* Devlin, K.*;* Feltwell, T.*;* Gentles, S.*;* Hamlin, N.*;* Holroyd, S.*;* Hornsby, T.*;* Jagels, K.*;* Barrell, B. G. *et al. Nature*, **1998**, *393*, 537.

[15] Ramakrishnan, L.*;* Federspiel, N. A.; Falkow, S. *Science*, **2000**, *288*, 1436.

[16] Goulding, C. W.*;* L. J. Perry, D.; erson, , M. R. Sawaya, D. C.*;* A. Parseghian, S. Chan? S. Nandy, V. C.*;* H. Gill; D. Eisenberg. *Trans. ACA*, **2001**, *36*, in press.

[17] Gill, H. S.*;* Pfluegl, G. M.; Eisenberg, D, *Acta Crystallogr. D*, **1999**, *55*, 865.

[18] Anderson, D. H.; Harth, G.; Horwitz, M. A.; Eisenberg, D. *J. Mol. Biol*, **2001**, *307*, 671.

[19] Harth, G.; Horwitz, M. A. *J Exp. Med*, **1999**, *189*, 1425.

[20] Tullius, M. V.; Harth, G.; Horwitz, M. A. *Infect. Immun*, **2001**, *69*, 6348.

[21] Harth, G.; Clemens, D. L.; Horwitz, M. A. *Proc. Natl. Acad. Sci. USA*, **1994**, *91*, 9342.

[22] Hirschfield, G. R.; McNeil, M.; Brennan, P. J. *J. Bacteriol*, **1990**, *172*, 1005.

[23] Gill, H. S.; Pfluegl, G. M. U.; Eisenberg, D, **2002**, manuscript in preparation.

[24] Gill, H. S.; Eisenberg, D. *Biochemistry*, **2001**, *40*, 1903.

[25] Belisle, J. T.; Vissa, V. D.; Sievert, T.; Takayama, K.; Brennan, P. J.; Besra, G. S. *Science*, **1997**, *276*, 1420.

[26] Horwitz, M. A.; Lee, B. W.; Dillon, B. J.; Harth, G. *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 1530.

[27] Ronning, D. R.; Klabunde, T.; Besra, G. S.; Vissa, V. D.; Belisle, J. T.; Sacchettini, J. C. *Nature Struct. Biol*, **2000**, *7*, 141.

[28] Nardini, M.; Ridder, I. S.; Rozeboom, H. J.; Kalk, K. H.; Rink, R.; Janssen, D. B.; Dijkstra, B. W. *J. Biol. Chem*, **1999**, *274*, 14579.

[29] Debarbieux, L.; Beckwith, J. *Cell*, **1999**, *99*, 117.

[30] Li, Q.; Hu, H.; Xu, G. *Biochem. Biophys. Res. Commun.* , **2001**, *283*, 849.

[31] Gomez, M.; Johnson, S.; Gennaro, M. L. *Infect. Immun*, **2000**, *68*, 2323.

[32] Wiker, H. G.; Michell, S. L.; Hewinson, R. G.; Spierings, E.; Nagai, S.; Harboe, M. *Microb. Pathog*, **1999**, *26*, 207.

[33] Martin, J. L. *Structure*, **1995**, *3*, 245.

[34] Goulding, C. W.; Apostol, M.; Parseghian, A.; Gennaro, M.; Eisenberg, D, **2002**, manuscript in preparation.

[35] Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 9367.

[36] Achari, A.; Somers, D. O.; Champness, J. N.; Bryant, P. K.; Rosemond, J.; Stammers, D. K. *Nature Struct. Biol*, **1997**, *4*, 490.

[37] Haussmann, C.; Rohdich, F.; Schmidt, E.; Bacher, A.; Richter, G. *J. Biol. Chem*, **1998**, *273*, 17418.

[38] Huovinen, P.; Sundstrom, L.; Swedberg, G.; Skold, O. *Antimicrob. Agents Chemother*, **1995**, *39*, 279.

[39] Locher, H. H.; Schlunegger, H.; Hartman, P. G.; Angehrn, P.; Then, R. L. *Antimicrob. Agents Chemother*, **1996**, *40*, 1376.

[40] Hennig, M.; D'Arcy, A.; Hampele, I. C.; Page, M. G.; Oefner, C.; Dale, G. E. *Nature Struct. Biol*, **1998**, *5*, 357.

[41] Ploom, T.; Haussmann, C.; Hof, P.; Steinbacher, S.; Bacher, A.; Richardson, J.; Huber, R. *Structure Fold. Des*, **1999**, *7*, 509.

[42] Goulding, C. W.; Apostol, M.; Parseghian, A.; Gennaro, M.; Eisenberg, D, 2002, manuscript in preparation.

[43] Manca, C.; Lyashchenko, K.; Wiker, H. G.; Usai, D.; Colangeli, R.; Gennaro, M. L.; *Infect. Immun*, **1997**, *65*, 16.

[44] Lee, B. Y.; Horwitz, M. A. *Infect. Immun*, **1999**, *67*, 2665.

[45] Braunstein, M.; Griffin, T. I.; Kriakov, J. I.; Friedman, S. T.; Grindley, N. D.; Jacobs, W. R. Jr. *J Bacteriol*, **2000**, *182*, 2732.

[46] Goulding, C. W.; Parseghian, A.; Apostol, M.; Sawaya, M.; Cascio, D.; Gennaro, M.; Eisenberg, D, **2002**, manuscript in preparation.

[47] Halaby, D. M.; Mornon, J. P. *J Mol. Evol*, **1998**, *46*, 389.

[48] Hirsch, J. A.; Schubert, C.; Gurevich, V. V.; Sigler, P. B. *Cell*, **1999**, *97*, 257.

[49] Owen, D. J.; Vallis, Y.; Pearse, B. M.; McMahon, H. T.; Evans, P. R. *EMBO J*, **2000**, *19*, 4216.

[50] Hamburger, Z. A.; Brown, M. S.; Isberg, R. R.; Bjorkman, P. J. *Science*, **1999**, *286*, 291.

[51] Gatfield, J.; Pieters, J. *Science*, **2000**, *288*, 1647.

[52] George, K. M.; Yuan, Y.; Sherman, D. R.; Barry, C. E.; 3rd, *J. Biol. Chem*, **1995**, *270*, 27292.

[53] Glickman, M. S.; Cox, J. S.; Jacobs, W. R. Jr. *Mol. Cell*, **2000**, *5*, 717.

[54] Yuan, Y.; Barry, C. E. 3rd. *Proc Natl Acad Sci U S A*, **1996**, *93*, 12828.

[55] Yuan, Y.; Lee, R. E.; Besra, G. S.; Belisle, J. T.; Barry, C. E.; 3rd, *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 6630.

[56] Wang, A. Y.; Grogan, D. W.; Cronan, J. E. Jr. *Biochemistry*, **1992**, *31*, 11020.

[57] Glickman, M. S.; Cahill, S. M.; Jacobs, W. R. Jr. *J. Biol. Chem*, **2001**, *276*, 2228.

[58] Minnikin, D. E.. In *The biology of mycobacteria*, **1982**, Academic Press, San Diego, pp. 95-184.

[59] Dubnau, E.; Chan, J.; Raynaud, C.; Mohan, V. P.; Laneelle, M. A.; Yu, K.; Quemard, A.; Smith, I.; Daffe, M. *Mol. Microbiol.* , **2000**, *36*, 630.

[60] Yuan, Y.; Crane, D. C.; Musser, J. M.; Sreevatsan, S.; Barry, C. E.; 3rd. *J Biol. Chem*, **1997**, *272*, 10041.

[61] Huang C. C.; Smith, C. V.; Glickman, M.; Jacobs, W. R. Jr.; Sacchettini, J. C. *J Biol. Chem*, **2001**, *26*, 26.

[62] Cheng, X.; Roberts, R. J. *Nucleic Acids Res.* , **2001**, *29*, 3784.

[63] Djordjevic, S.; Stock, A. M. *Structure*, **1997**, *5*, 545.

[64] Klimasauskas, S.*;* Kumar, S.*;* Roberts, R. J.; Cheng, X. *Cell*, **1994**, *76*, 357.

[65] Schluckebier, G.*;* O'Gara, M.*;* Saenger, W.; Cheng, X. *J. Mol. Biol*, **1995**, *247*, 16.

[66] Parris, K. D.*;* Lin, L.*;* Tam, A.*;* Mathew, R.*;* Hixon, J.*;* Stahl, M.*;* Fritz, C. C.*;* Seehra, J.; Somers, W. S. *Structure Fold. Des*, **2000**, *8*, 883.

[67] Zhang, Y. M.*;* Rao, M. S.*;* Heath, R. J.*;* Price, A. C.*;* Olson, A. J.; Rock, C. O.; White, S. W. *J Biol. Chem*, **2001**, *276*, 8231.

[68] McKinney, J. D.*;* Honer zu Bentrup, K, , Munoz-Elias, E. J.*;* Miczak, A, , Chen, B.*;* Chan, W. T.*;* Swenson, D.*;* Sacchettini, J. C.*;* Jacobs, W. R. Jr.; Russell, D. G. *Nature*, **2000**, *406*, 735.

[69] Sharma, V.*;* Sharma, S.*;* Hoener zu Bentrup, K.*;* McKinney, J. D.*;* Russell, D. G.*;* Jacobs, W. R.; Sacchettini, J. C. *Nature Struct. Biol*, **2000**, *7*, 663.

[70] Blanchard, J. S. *Annu. Rev. Biochem*, **1996**, *65*, 215.

[71] Banerjee, A.; Dubnau, E.; Quemard, A.; Balasubramanian, V.; Um, K. S.; Wilson, T.; Collins, D.; Delisle, G.; Jacobs, W. R. Jr. *Science*, **1994**, *263*, 227.

[72] Vilcheze C.*;* Morbidoni, H. R.*;* Weisbrod, T. R.*;* Iwamoto, H.*;* Kuo, M.*;* Sacchettini, J. C.*;* Jacobs, W. R. *J. Bacteriol.* **2000**, *182*, 4059.

[73] Baldock, C.; Rafferty, J. B.; Sedelnikova, S. E. ? Baker, P. J.; Stuitje, A. R.; Slabas, A. R.; Hawkes, T. R.; Rice, D. W. *Science*, **1996**, *274*, 2107.

[74] Rafferty, J. B.*;* Simon, J. W.*;* Baldock, C.*;* Artymiuk, P. J.*;* Baker, P. J.*;* Stuitje, A. R.*;* Slabas, A. R.; Rice, D. W. *Structure*, **1995**, *3*, 927.

[75] Heath, R. J.*;* Yu, Y. T.*;* Shapiro, M. A.*;* Olson, E.*;* & Rock, C. O. *J. Biol. Chem*, **1998**, *273*, 30316.

[76] Ward, W. H. J.*;* Holdgate, G. A.*;* Rowsell, S.*;* McLean, E. G.*;* Pauptit, R. A.*;* Clayton, E.*;* Nichols, W. W.*;* Colls, J. G.*;* Minshull, C. A.*;* Jude, D. A.*;* Mistry, A. *Biochemistry*, **1999**, *38*, 12514.

[77] Heath, R. J.*;* Rubin, J. R.*;* Holland, D. R.*;* Zhang, E. L.*;* Snow, M. E. & Roch, C. O. *J. Biol. Chem*, **1999**, *274*, 11110.

[78] Stewart, M. J.; Parikh, S.; Xiao, G. P.; Tonge, P. J.; Kisker, C. *J. Mol. Biol*, **1999**, *290*, 859.

[79] Levy, C. W.*;* Roujeinikova, A.*;* Sedelnikova, S.*;* Baker, P. J.*;* Stuitje, A. R.*;* Rice, D. W.; Rafferty, J. B. *Nature*, **1999**, *398*, 383.

[80] Roujeinikova, A.*;* Sedelnikova S.*;* deBoer, G. J.*;* Stuitje, A. R.*;* Slabas, A. R.*;* Rafferty, J. B.*;* Rice, D. W. *J. Biol. Chem*, **1999**, *294*, 527.

[81] Parikh, S. L.; Xiao, G. P.; Tonge, P. J. *Biochemistry*, **2000**, *39*, 7645.

[82] Banerjee, A.*;* Sugantino, M.*;* Sacchettini, J. C.; Jacobs, W. R Jr. *Microbiology*, **1998**, *144*, 2697.

[83] Waldo, G. S.; Standish, B. M.; Berendzen, J.; Terwilliger, T. C. *Nature Biotechnol.*, **1999**, *17*, 691.

[84] Yang, J. K.; Yoon, H.-J.; Ahn, H. J.; Lee, B. I.; Cho, S. H.; Waldo, G. S.; Park, M. S.; Suh, S. W. *Acta Crystallographica D*, **2002**, *58*, 303.

[85] Kozloff, L. M.; Turner, M. A.; Arellano, F;Lute, M. *J. Bacteriol*, **1991**, *173*, 2053.

[86] Chen, I.;Charalompous, F. C. *J. Biol. Chem*, **1964**, *239*, 1905.

[87] Chen, I.;Charalompous, F. C. *Biochem. Biophys. Res. Commun*, **1965**, *19*, 144.

[88] Culbertson, M. R.;Henry, S. A. *Genetics*, **1975**, *80*, 23.

[89] Bachhawat, N.; Mande, S. C. *J. Mol. Biol*, **1999**, *291*, 531.

[90] Bachhawat, N.; Mande, S. C. *Trends Genet.* , **2000**, *16*, 111.

[91] Altschul, S. F.*;* Madden, T. L.*;* Schaffer, A. A.*;* Zhang, J. H.*;* Zhang, Z.*;* Miller, W.*;* Lipman, D. J *Nucl. Acids Res.* , **1997**, *25*, 3389.

[92] Hofmann, K.; Bucher, P.; Falquet, L.;; Bairoch, A. *Nucl. Acids Res.* **1999**, *27*, 215.

[93] Greenberg, M. L.; Reiner, B.;Henry, S. A. *Genetics*, **1982**, *100*, 19.

[94] Sherman, D. R.*;* Sabo, P. J.*;* Hickey, M. J.*;* Arain, T. M.*;* Mahairas, G. G.*;* Yuan, Y.*;* Barry, 3rd C. E.; Stover, C. K. *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 6625.

[95] Wilson, M.*;* Derisi, J.*;* Kristenson, H. H.*;* Imboden, P.*;* Rane, S.; Brown, P. O. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 12833.

[96] Wilson, T.*;* de Lisle, G. W.*;* Marcinkeviciene, J. A.*;* Blanchard, J. S.; Collins, D. M. *Mol. Microbiol.* , **1998**, *144*, 2687.

[97] Ellis, H. R.;Poole, L. B. *Biochemistry*, **1997a**, *36*, 13349.

[98] Ellis, H. R.;Poole, L. B. *Biochemistry*, **1997b**, *36*, 15013.

[99] Chauhan, R.;Mande, S. C. *Biochem. J*, **2001**, *354*, 209.

[100] Chauhan, R;Mande, S. C, **2002**, Manuscript submitted for publication.

[101] Hillas, P. J.; del Alba, F. S.; Oyarzabal, J.; Wilks, A.;Ortiz De Montellano, P. R. *J. Biol. Chem*, **2000**, *275*, 18801.

[102] Schröder, E.; Littlechild, J. A.; Lebedev, A. A.; Errington, N.; Vagin, A. A.;Isupov, M. N. *Structure*, **2000**, *8*, 605.

[103] Meganathan R. *Vitamins, Hormones*, **2001**, *61*, 173.

[104] Meganathan R. in Eschericha coli, Salmonella typhimurium. Cellular, Molecular Biology Volume 1, **1996**, F. C. Neidhardt Ed. , pp. 642-656.

[105] Dibrov, E.*;* Robinson, K. M.; Lemire, B. D. *J. Biol. Chem*, **1997**, *272*, 9175.

[106] Lee P. T.*;* Hsu A. Y.*;* Ha H. T.; Clarke C. F. *J. Bacteriol*, **1997**, *179*, 1748.

[107] Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. *J. Mol. Biol*, **1995**, *247*, 536.

[108] Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M. *Structure*, **1997**, *5*, 1093.

[109] Gupta, A.; Kumar, P. H.; Dineshkumar, T. K.; Varshney, U. & Subramanya, H. S. *J. Mol. Biol*, **2001**, *312*, 381.

[110] Holm, L.; Sander, C. *Science*, **1996**, *273*, 595.

[111] Herzberg O.; Chen C. C. H.; Kapadia G.; McGuire M.; Carroll L. J.; Noh S. J.; Dunaway-Mariano, D. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*, 2652.

[112] Sligar, S. G.; Gunsalus, I. C. *Proc. Natl. Acad. Sci. USA*, **1976**, *73*, 1078.

[113] Nelson, D. R.; Koymans, L.; Kamataki, T.; Stegeman, J. J.; Feyereisen, R.; Waxman, D. J.; Waterman, M. R.; Gotoh, O.; Coon, M. J.; Estabrook, R. W.; Gunsalus, I. C.; Nebert, D. W. *Pharmacogenetics*, **1996**, *6*, 1.

[114] Bellamine, A.; Mangla, A. T.; Nes, W. D.; Waterman, M. R. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 8937.

[115] Souter, A.; McLean, K. J.; Smith, W. E.; Munro, A. W. *J. Chem. Technol. Biotechnol*, **2000**, *75*, 933.

[116] CuppVickery, J. R.; Han, O.; Hutchinson, C. R.; Poulos T. L. *Nature Struct. Biol*, **1996**, *3*, 632.

[117] Mercer, E. I. *Lipids*, **1991**, *26*, 584.

[118] Vidakovic, M.; Sligar, S. G.; Li, H. Y.; Poulos, T. L. *Biochemistry*, **1998**, *37*, 9211.